Identification Of Regulatory Elements Within The Rat Preprotachykinin-A Promoter

by

Susan C. Mendelson

Thesis submitted for the Degree of Doctor of Philosophy

University of Edinburgh
April 1994

This thesis is dedicated to my parents whose moral and financial support have made this possible.

Acknowledgements

I would like to thank the Medical Research Council for the award of a research studentship and Professor George Fink for the opportunity to study in the MRC Brain Metabolism Unit.

Thanks also go to John Quinn for his supervision throughout my period of research in the unit and to both John Quinn, Tony Harmar and George Fink for their careful and critical reading of this thesis. I am also grateful to John Sheward for his help with making figures and to Frank and Graham for the excellent photography service they provided throughout this research project.

I would like to thank everyone at the Royal Edinburgh Hospital for their friendship and support during the last three and a half years, Christine and Elma for their help with tissue culture work and Janice for her advice and support when the going got tough.

Finally, I would like to thank my parents for their financial and moral support throughout my time at University, my sister Jane for support and advice during the hard times and most importantly John Morrow for his endless support, patience, love and regular trips to the Canny Mans which helped keep me sane during the long months of writing up.

Publications Arising From This Thesis

Mendelson, S.C., C. Morrison and J.P. Quinn (1992). An NGF-inducible octamer binding protein activity in a C1300 neuroblastoma cell line. Molecular Brain Research, 15, 174-178.

Mendelson, S.C., and J.P. Quinn (1992). Identification of a neuronal specific DNA-binding protein within the rat preprotachykinin promoter. Biochemical Society Transactions, 20, 264S.

Mendelson, S.C., and J.P. Quinn (1993). Identification of potential regulatory elements within the rat pre-protachykinin A promoter. Biochemical Society Transactions, 21, 372S.

McAllister, J., S. Mendelson, and J.P. Quinn (1993). The preprotachykinin A promoter interacts with a sequence specific single stranded DNA binding protein. Biochemical Society Transactions, 21, 373S.

Quinn, J.P., C. Morrison, J. McAllister, and S. Mendelson (1993). Evolution of enhancer domains within the preprotachykinin promoter. Biochemical Society Transactions, 21, 371S.

Abstract

The preprotachykinin-A (PPT-A) gene encodes the neuropeptides substance P, neurokinin A, neuropeptide K and neuropeptide γ, each derived by alternative mRNA splicing and post-translational processing of precursor polypeptides.

This study set out to characterise the cis-acting elements which control expression of the rat PPT-A (rPPT-A) gene through analysis of 1300 bp of promoter sequence spanning nucleotides -865 to +447 relative to the transcriptional start site. By the use of exonuclease protection assays, DNase1 footprinting and electrophoretic mobility shift analysis several potential regulatory elements have been identified both 5' and 3' of the major transcriptional start site.

The rPPT-A gene is expressed endogenously, at high levels, in dorsal root ganglion (DRG) neurons, however this tissue cannot be obtained in sufficient quantity for biochemical analysis. To date, there are no cell lines which express the rPPT-A gene at high levels, either endogenously or when transfected with plasmids containing regions of the promoter linked to reporter genes. Therefore, the well characterised HeLa cell line was used as a source of protein extract as it has been shown to express a large array of transcription factors, many of which are expressed in DRG neurons.

By use of the exonuclease protection assay a protein/DNA interaction site was identified within the rPPT-A promoter, close to the transcriptional start site. In order to map this and any other sites more accurately, DNase1 footprinting analysis was carried out and identified elements were then further characterised by electrophoretic mobility shift assays. Several potential regulatory elements which contain the potential binding sites for a number of distinct transcription factors were identified . These include four AP-1-like sites, one of which also resembles a cyclic AMP response element (CRE), a purine-rich element and six E-box motifs. There are also two regions rich in dT nucleotides, resembling octamer binding protein motifs, and two regions rich in dG nucleotides. Other elements include one located between the TATA box and transcriptional site, another which appears to bind both single-stranded and double-stranded DNA-binding proteins in a tissue specific manner and two elements which have yet to be further characterised. Some of the elements located 3' of the major transcriptional start site differ in their binding characteristics depending on whether extracts of neuronal or non-neuronal tissue were used.

The rPPT-A promoter is therefore bound by a variety of proteins, some of which may bind in a tissue-specific manner, thus potentially allowing for stimulus-dependent and tissue-specific gene regulation. It is hoped that the sites identified by this study will act as a foundation for studies of the regulation of the rPPT-A promoter in DRG neurons.

In a complementary study it was demonstrated, by electrophoretic mobility shift assays using the mouse C1300 neuroblastoma cell line, that nerve growth factor (NGF) induces a protein, termed C1300-Oct, which recognises the octamer binding protein consensus sequence. As NGF has been shown to regulate the rPPT-A gene in DRG neurons and the rPPT-A promoter contains at least two potential octamer binding protein sites, it is postulated that NGF-inducible octamer binding proteins, including C1300-Oct, may be involved in the regulation of rPPT-A gene expression.

Abbreviations

A	adenine
AAD	acidic activation domain
Ac	acetate
AdML	adenovirus major late
APS	ammonium persulphate
ATP	adenosine-5'-triphosphate
bHLH	basic helix-loop-helix
bp	base pair(s)
BSA	bovine serum albumin
bZIP	basic leucine zipper
C	cytosine
Ci	curies
CIP	calf intestinal phosphatase
cm	centimetre
CNS	central nervous system
CTD	COOH-terminal domain
Cys	cysteine
Da	daltons
dA	2'-deoxyadenosine
dATP	2'-deoxyadenosine-5'-triphosphate
dC	2'-deoxycytidine
dCTP	2'-deoxycytidine-5'-triphosphate
ddNTP	2', 3'-dideoxyribonucleoside 5'-triphosphate
dG	2'-deoxyguanosine
dGTP	2'-deoxyguanosine-5'-triphosphate
dH_2O	distilled water
dT	2'-deoxythymidine
dTTP	2'-deoxythymidine-5'-triphosphate
dNTP	2'-deoxyribonucleoside-5'-triphosphate
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DRG	dorsal root ganglion (ganglia)
DTT	dithiothreitiol
E.coli	*Escherichia coli*
EDTA	sodium ethylenediamine tetra-acetic acid
EtBr	ethidium bromide

g	gram(s)
G	guanine
Gly	glycine
h	hour (s)
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid
His	histidine
HMG	high mobility group
HSH	helix-span-helix
HSV	herpes simplex virus
HTH	helix-turn-helix
IE	immediate early (gene)
IPTG	isopropylthio-β-D-galactoside
kb	kilobase pairs
KDa	kilo daltons
l	litre(s)
Leu	leucine
m	metre(s)
M	molar
mA	milliamp(s)
Met	methionine
min	minute(s)
ml	millilitre(s)
mm	millimetre(s)
mM	millimolar
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger ribonucleic acid
M_r	molecular weight
N	unspecified nucleotide or amino acid
n	nano
NGF	nerve growth factor
NKA	neurokinin A
NKB	neurokinin B
Npγ	neuropeptide γ
NpK	neuropeptide K
OD	optical density
^{32}P	phosphorous 32 radioisotope
PBS	phosphate buffered saline
PCR	polymerase chain reaction

Phe	phenylalanine
PMSF	phenylemethylsulfonylflouride
PPT	preprotachykinin
Pro	proline
Py	pyrimidine moiety
R	purine moiety
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
^{35}S	sulphur 35 radioisotope
SDS	sodium dodecyl sulphate
sec	second(s)
Ser	serine
SP	substance P
SV40	simian virus 40
T	thymidine
TAF	TBP-associated factors
TBP	TATA-box binding protein
TBE	tris-borate/EDTA
TEMED	N,N,N',N'- teramethylethylene diamine
Thr	threonine
Tris	tris (hydroxymethyl) aminoethane
tRNA	transfer ribonucleic acid
Tyr	tyrosine
UAS	upstream activating sequence
UV	ultraviolet
V	volt
vol	volume
v/v	volume/volume
w/v	weight/volume
X	hydrophobic or aromatic residue
X-gal	5-bromo-4-chloro-3-indolyl-B-D-galactopyranoside
Y	pyrimidine moiety

Greek Symbols

α	alpha
β	beta
δ	delta
γ	gamma
μ	micro

Contents.

Section 1: Introduction.

1.1. Tachykinins.

This study set out to investigate how the rat preprotachykinin-A (rPPT-A) gene is regulated at the transcriptional level.

The rPPT-A gene encodes the precursors of the tachykinins (literally 'fast-acting') which are a family of neuropeptides displaying a wide but selective distribution throughout both the central and peripheral nervous systems. The peptides vary in length but share substantial amino acid sequence homology and are characterised by the common COOH-terminal amino acid sequence, essential for their biological activity, Phe-X-Gly-Leu-Met-NH_2 where X is a hydrophobic or an aromatic residue (Erspamer, 1971).

The tachykinin Substance P (SP) was first purified by Chang *et al.* (1971) and its structure was elucidated as the undecapeptide Arg-Pro-Lys-Pro-Gln-Gln-Phe-Phe-Gly-Leu-Met-NH_2. SP was originally the only representative tachykinin in mammals, however further mammalian tachykinins have subsequently been described. These include neurokinin A (NKA) (Nawa *et al.*, 1984; Krause *et al.*, 1987), and the NH_2-terminal extended derivatives of NKA, NKA (3-10) (Tatemoto *et al.*, 1985), neuropeptide K (NpK) (Tatemoto *et al.*, 1985) and neuropeptide γ (Npγ) (Kage *et al.*, 1988) and finally neurokinin B (NKB) (Kotani *et al.*, 1986; Bonner *et al.*, 1987).

1.1.1. Genes encoding tachykinins.

In mammals the tachykinin peptides are derived by post-translational processing of polyprotein preprotachykinin (PPT) precursors. There are two genes that encode the tachykinin peptides. The PPT-A gene (or SP/NKA gene) encodes SP, NKA, NKA (3-10), NpK and Npγ (Nawa *et al.*, 1983, 1984; Krause et al., 1987) and the PPT-B gene encodes NKB (Kotani *et al.*, 1986; Bonner *et al.*, 1987).

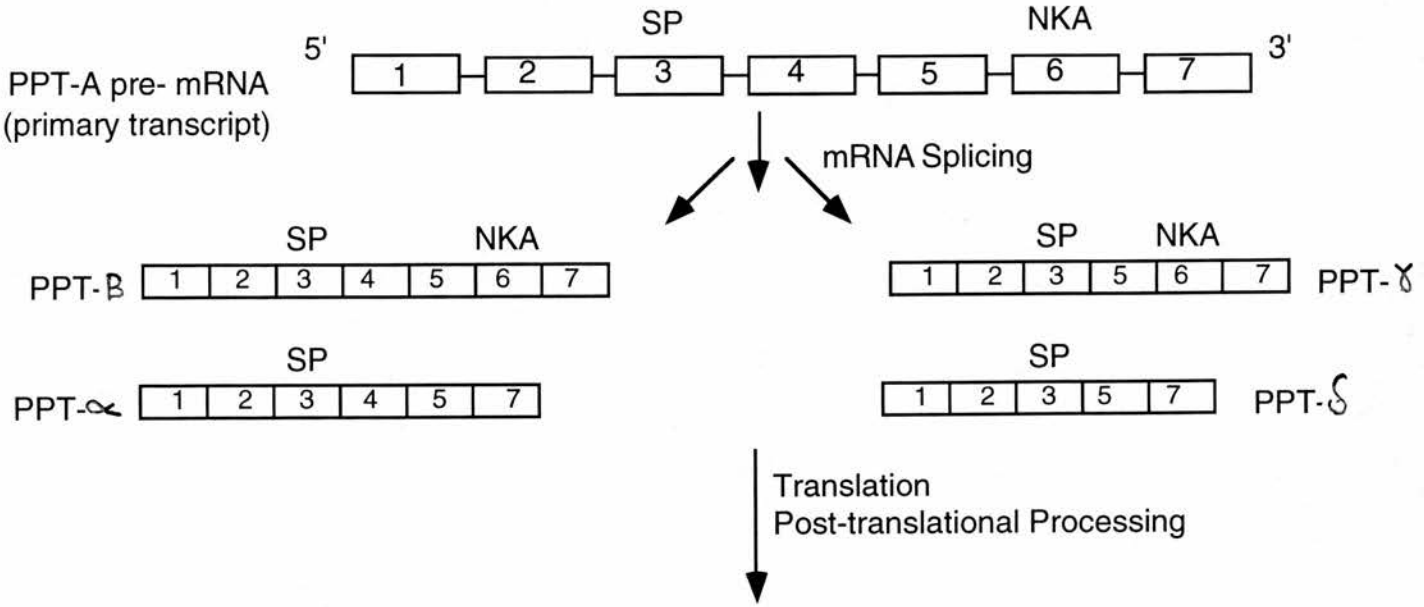
The PPT-A gene is approximately 8.4 Kbp long and consists of seven exons interspersed by six introns (Figure 1). Sequences in exon 3 encode SP and sequences in exon 6 encode NKA. The coding region of the PPT-A gene consists of 123 bp of the 3' end of exon 2, exons 3-6 and 47 bp of the 5' end of exon 7. Alternative mRNA splicing of the primary transcript results in the generation of four mRNAs termed α-, β-, γ-, and δ-PPT-A mRNA (Nawa *et al.*, 1983, 1984; Kawaguchi *et al.*, 1986; Krause *et al.*, 1987; Harmar *et al.*, 1990). Exons 6 and 4 are separately excluded in the splicing process resulting in the formation of α-PPT-A and γ-PPT-A mRNA respectively, β-PPT-A mRNA contains regions from all exons (Nawa *et al.*, 1983, 1984; Kawaguchi *et al.*, 1986; Krause *et al.*, 1987) and δ-PPT-A mRNA lacks both exon 4 and exon 6 (Harmar *et al.*, 1990).

Figure 1. Schematic diagram of the rat preprotachykinin-A gene.

The primary transcript of the rPPT-A gene is depicted at the top. The boxes, numbered 1 to 7, denote exons and are separated by introns, depicted as lines. Sequences in exon 3 encode substance P (SP) and sequences in exon 6 encode neurokinin A (NKA).

The primary transcript undergoes alternate mRNA splicing to generate four mRNA species: α, β, γ, and δ.

The polypeptide precursors encoded by these mRNA species then undergo post-translational processing to generate the neuropeptides SP, NKA and the NH_2-terminal extended derivatives of NKA, NKA (3-10) neuropeptide K (NpK) and neuropeptide γ (Npγ).



SP (all 4 mRNAs)

NKA (PPT-β and PPT-γ)

NKA (3-10) (PPT-β and PPT-γ)

NpK (PPT-β)

Npγ (PPT-γ)

Differential post-translational processing of the PPT-A mRNAs results in the production of SP from all four PPT-A mRNAs, NKA, NKA (3-10) and/or NpK from β-PPT-A mRNA (Tatemoto *et al.*, 1985; MacDonald *et al.*, 1989) and NKA, NKA (3-10) and/or Npγ from γ-PPT-A mRNA (Kage *et al.*, 1988; MacDonald *et al.*, 1989). δ-PPT-A mRNA is predicted, in addition to SP, to encode a novel COOH-terminal peptide of 22 amino acids unique to the δ-PPT transcript (Harmar *et al.*, 1990).

Tissue-and species-specific patterns of splicing of the primary transcript have been reported for the bovine and the rat PPT-A genes. For example, Northernblot and *in situ* hybridisation studies have shown that bovine β-PPT-A mRNA is more abundant in CNS tissue compared to α-PPT-A mRNA, whereas α-PPT-A mRNA predominates in thyroid and gut (Nawa *et al.*, 1984). In rat tissues γ-PPT-A mRNA was found to be the most abundant form with the splicing pattern of the primary transcript being relatively constant in all tissues studied (Krause *et al.*, 1987, 1989; Carter and Krause, 1990). To date, δ-PPT-A mRNA has only been reported in rat dorsal root ganglia (DRG) (Harmar *et al.*, 1990).

In an analogous situation to the PPT-A gene, the PPT-B gene consists of 7 exons interspaced with six introns with sequences encoding NKB contained in exon 5. The PPT-B gene gives rise to two PPT-B mRNAs through the use of two different promoter sites. Both mRNA species contain exons 1-7 and encode a propeptide of 126 amino acids which is processed to produce NKB, the sole biologically active peptide of the PPT-B gene (Kotani *et al.*, 1986; Bonner *et al.*, 1987).

1.1.2. Distribution of tachykinins.

A number of techniques have been used to study the distribution of PPT-A mRNAs and tachykinin peptides in the CNS and other tissues. Such techniques include radioimmunoassays combined with chromatographic techniques (Ogawa *et al.*, 1985; Arai and Emson, 1986; Too *et al.*, 1989), PCR, *in situ* hybridisation, nuclease protection and Northern blot assays (Nawa *et al.*, 1983, 1984; Krause *et al.*, 1987, 1989; Carter and Krause, 1990, Harmar *et al.*, 1990; Bannon *et al.*, 1992). High levels of SP, NKA and NpK immunoreactivity have been detected in the spinal cord and in sensory neurons of the DRG (Ogawa *et al.*, 1985; Arai and Emson, 1986; Too *et al.*, 1989). PPT-A mRNAs have been detected in a wide range of CNS tissues including the spinal cord, DRG, hippocampus, cerebellum, striatum and basal ganglia (Nawa *et al.*, 1983, 1984; Krause *et al.*, 1987, 1989; Carter and Krause, 1990; Harmar *et al.*, 1990; Bannon *et al.*, 1992).

Northern blot, *in situ* hybridisation and nuclease protection assays have also shown that the different PPT-A mRNAs are expressed in a variety of peripheral tissues

including the intestine, the stomach, the thyroid and the retina (Nawa *et al.*, 1983, 1984; Krause *et al.*, 1987, 1989; Carter and Krause, 1990).

1.1.3. Central roles of tachykinins.

Substance P (SP) is found in discrete neuronal pathways of the central nervous system and is thought to act as a neurotransmitter in primary sensory neurons (Otsuka *et al.*, 1982).

Early evidence to support the role of SP as a neurotransmitter came from studies concerning the distribution of SP in the spinal cord. Such studies demonstrated that SP is more abundant in the sensory neurons of the dorsal horn, where DRG neurons terminate, than in the ventral horn (Lembeck, 1953). Subsequently it was shown that SP is highly concentrated in the nerve terminals of DRG sensory neurons (Takahashi and Otsuka, 1975). Immunohistochemistry later confirmed these findings and allowed the localisation of SP to nerve terminals forming synapses with dorsal horn cells (Barber *et al.*, 1979).

Further evidence supporting a neurotransmitter role for SP comes from the finding that SP, extracted from frog dorsal roots, is a powerful stimulant of spinal neurons (Otsuka *et al.*, 1972). Additionally, when the isolated dorsal roots of the spinal cords of new born rats are stimulated there is a marked increase in the amount of immunoreactive SP in the solution perfusing the preparation (Otsuka and Konishi, 1976; Akagi *et al.*, 1980; Yaksh *et al.*, 1980).

The presence of other tachykinins, such as NKA and NpK, have also been demonstrated in primary sensory neurons (Minamino *et al.*, 1984; Dalsgaard *et al.*, 1985; Hua *et al.*, 1985; Ogawa *et al.*, 1985). Additionally, SP and NKA are co-released from nerve terminals (Saria *et al.*, 1984) and have been shown to have similar actions on effector cells (Hua *et al.*, 1985; Nilsson *et al.*, 1985).

DRG neurons are believed to be mediators of pain perception and a number of lines of evidence point to SP as a candidate factor responsible for mediating the transmission of nociceptive information to the spinal cord. Such evidence includes the finding that capsaicin, which causes a desensitisation to chemical noxious stimuli, evokes the release of SP and in fact depletes SP, from dorsal roots (Gasparovic *et al.*, 1964; Jessel *et al.*, 1978; Nagy *et al.*, 1981). Additionally, iontophoretic application of SP onto the dorsal horn exerts a powerful excitatory action on a subpopulation of dorsal horn neurons which are also excited by a peripheral noxious stimulus (Henry, 1976; Randic and Miletic, 1977). SP has also been found to be released during noxious mechanical or thermal stimulation (Kuraishi *et al.*, 1985; Wiesenfeld-Hallin, 1986) and noxious stimulation has been found to lead to the induction of the PPT-A gene in rat DRG (Noguchi *et al.*, 1988). Furthermore, the number of dorsal horn projection

neurons in rats expressing PPT-A mRNA was found to increase following noxious stimulus (Noguchi and Ruda, 1992), providing further evidence that tachykinin peptides act as neurotransmitters in nociceptive CNS pathways.

1.1.4. Peripheral roles of tachykinins.

The majority of SP synthesised in the dorsal roots is transported to the periphery (Brimijion *et al.*, 1980; Harmar and Keen, 1982). Consequently, SP has been demonstrated in a variety of peripheral nerves, including phrenic (Malthe-Sorensen and Oktedalon, 1982) and sciatic nerves (Lundberg *et al.*, 1978; Brimijoin *et al.*, 1980; Lembeck *et al.*, 1981). Additionally, SP-containing nerve fibers have been demonstrated in the skin (Hokfelt *et al.*, 1975; Dalsgaard *et al.*, 1983), around blood vessels (Furness *et al.*, 1982), in the rat mammary nipple (Traurig *et al.*, 1984), in different areas of the eye (Keen *et al.*, 1982; Dalsgaard, 1988), in the spleen, thymus and lymph nodes (Felten *et al.*, 1985; Fink and Weihe, 1988; Popper *et al.*, 1988; Weihe *et al.*, 1989; Kurkowski *et al.*, 1990) and in the intestine (Bienenstock *et al.*, 1987).

In the cornea, capsaicin treatment results in a depletion of SP (Keen *et al.*, 1982) and capsaicin treatment of blood vessels results in a depletion of SP from innervating nerve fibers (Furness *et al.*, 1982; Papka *et al.*, 1984).

In addition to mediating the pathogenesis of pain, peripheral SP is also thought to be involved in processes such as neurogenic inflammation associated with arthritis (Pernow, 1985; Donaldson *et al.*, 1992; Smith *et al.*, 1993), stimulating immune function (Bar-Shavit *et al.*, 1980), hypotension, smooth muscle contraction and cellular proliferation (Bury and Mashford, 1976; Couture *et al.*, 1979).

1.1.5. Regulators of tachykinins.

The ability of the nervous system to respond to different types of stimuli depends on an extensive repertoire of adaptive responses, one of which is a change in the biosynthesis of neurotransmitter molecules.

The tachykinins and their precursor mRNAs have been shown to display a significant degree of plasticity in their expression levels. A number of factors have been identified as potential regulators of the tachykinins. One such regulator is the neurotrophic factor nerve growth factor (NGF). NGF has been shown to regulate the expression of PPT-A mRNA in adult rat DRG neurons (Lindsay and Harmar, 1989). DRG neurons require NGF for survival during development, but once mature they appear to survive in the absence of NGF (Johnson *et al.*, 1986). NGF deprivation in adults rats, achieved through sciatic nerve section, results in a reduction of substance P and PPT-A mRNA levels in lumbar DRG and this effect can be reversed by the

application of NGF to the nerve (Fitzgerald *et al.*, 1985; Henken *et al.*, 1990). Evidence that the effects of added NGF are due to direct stimulation of SP biosynthesis came from the findings of Lindsay and Harmar (1989). Adult DRG neurons grown in the absence of NGF were shown by Northernblot and radioimmunoassays to display a reduced level of SP and its precursor mRNA, whereas cultures grown in the presence of NGF displayed an increase in PPT-A mRNA and SP levels over a time course of five days (Lindsay and Harmar, 1989).

Deafferentation of rat neonatal superior cervical ganglion (SCG) results in increased levels of SP and its mRNA (Kessler *et al.*, 1981; Roach *et al.*, 1987). The mechanisms underlying this increase are thought to involve immune cytokines which are released during such injury. It has been shown that interleukin-1β (IL-1β) results in a dramatic increase in SP-like immunoreactivity and in PPT-A mRNA in cultured neonatal SCG (Jonakait and Schotland, 1990; Jonakait *et al.*, 1990; Freidin and Kessler, 1991; Hart *et al.*, 1991). This response was shown to require the presence of non-neuronal cells (Freidin and Kessler, 1991). Subsequently, it was demonstrated that leukaemia inhibitory factor (LIF), which also promotes SP expression in SCG cultures (Nawa and Sah, 1990; Freidin and Kessler, 1991; Rao *et al.*, 1992), acts as an intermediate molecule stimulating cells to produce SP (Shadiack *et al.*, 1993). It appears that IL-1β does not directly act on neurons but rather that LIF, produced by IL-1β stimulated non-neuronal cells, acts as at least one of the molecular intermediates of IL-1β to promote SP expression (Shadiack *et al.*, 1993).

Brain tachykinins are functionally closely linked to dopaminergic transmission in the mammalian brain (Bjorklund and Lindvall, 1984). It has been shown that, in the basal ganglia, factors which modulate dopaminergic transmission also influence the synthesis and *in vivo* release of tachykinins (Ritter *et al.*, 1984; Bannon *et al.*, 1986, 1987; Lindefors *et al.*, 1989; Lindefors, 1992). Haloperidol, a dopamine receptor antagonist, has been shown to inhibit PPT-A mRNA expression in the basal ganglia (Bannon *et al.*, 1986, 1987; Lindefors *et al.*, 1989; Lindefors, 1992) and amphetamine, an indirect dopamine agonist, induces PPT-A mRNA expression (Ritter *et al.*, 1984; Bannon *et al.*, 1987; Lindefors, 1992). Dopaminergic neurons of the midbrain are also thought to regulate PPT-A mRNA expression as it has been shown that there is a down-regulation of PPT-A mRNA expression by a specific set of dopaminergic neurons in the midbrain during rat brain development (Brene *et al.*, 1992).

1.1.6. Tachykinin receptors.

SP and other tachykinins mediate their actions through G-protein linked receptors. A large number of G-protein-linked receptors exist, characterised as sharing a common structure of seven transmembrane (TM) domains, including the well

characterised β-adrenergic receptor and the visual pigment rhodopsin. Tachykinins, on binding to their receptors, are thought to activate phospholipase C (PLC). This results in the stimulation of phosphatidyl inositol turnover and a resulting increase in intracellular inositol 1,4,5-triphosphate (IP_3) which evokes Ca^{2+} release from intracellular stores. PLC activation also results in protein kinase C (PKC) activation through diacylglycerol (DAG) (Womack *et al.*, 1988; reviewed in Zeiglgansberger and Tolle, 1993).

All of the tachykinins are able to induce similar biological actions, although their rank order of potency at inducing individual physiological responses differs, suggesting that they may act through different receptors. Subsequently, ligand binding studies and molecular biological techniques have identified three tachykinin receptors, NK1, NK2 and NK3, preferentially bound by SP, NKA and NKB respectively (Mizrahi *et al.*, 1985; Buck and Burcher, 1986; Masu *et al.*, 1987; Yokota *et al.*, 1989; Shigemoto *et al.*, 1990).

1.2. Eukaryotic RNA polymerase II transcription.

1.2.1. An overview.

Many of the factors which have been shown to be potential regulators of rPPT-A gene expression may exert their actions through the regulation of the initiation of mRNA synthesis. A number of distinct DNA sequence elements, usually located upstream of the start point of transcription, are involved in regulating the synthesis of mRNA. These include minimal promoter elements, upstream activating sequences (UASs) and enhancers.

Minimal promoter elements commonly include a TATA box motif, located around -30 from the site of transcription initiation, and sometimes also include an initiator element(s). Such minimal promoter elements control the site of transcription initiation and cellular factors exist with which they will functionally interact.

UASs and enhancers can act to regulate the initiation of transcription by RNA polymerase II from the minimal promoter. UASs can act at a short distance while enhancer elements can be located up to several thousand base pairs from the transcription initiation site (Dynan and Tjian, 1985). These types of elements will be discussed in more detail in Section 1.2.3.

The control of transcription initiation by RNA polymerase II requires the co-ordinated interaction of transcription factors with different types of cis-acting DNA sequence elements. Two classes of transcription factors exist. Firstly, the general transcription factors, which will be covered in the following sections, are essential for

initiation and are sufficient to direct a basal level of transcription from many minimal promoters. Secondly, trans-acting DNA binding proteins, which will be covered in Section 1.2.3., bind to upstream activating sequences and enhancers to regulate the frequency of transcriptional initiation (Dynan and Tjian, 1985).

1.2.2. Minimal promoter.

The initiation of transcription by RNA polymerase II is a highly complex process and requires the action of multiple initiation factors. Transcriptionally active extracts were first prepared and fractionated from crude cellular extracts over 10 years ago (Weil *et al.*, 1979; Matsui *et al.*, 1980). Subsequent studies have resulted in the purification of many of the general transcription factors and in the cloning of the genes encoding these factors.

The following sections will give a general overview of the initiation of transcription by RNA polymerase II from the minimal promoter regions of TATA-box containing promoters. This will include a section covering the properties of RNA polymerase II itself, a section describing the general transcription factors and how they are thought to interact with each other to initiate transcription and a section covering initiator elements.

1.2.2.1. RNA polymerase II.

Eukaryotic cells contain three distinct forms of nuclear DNA-dependent RNA polymerases that transcribe different sets of genes. RNA polymerase I is responsible for synthesising rRNA precursors (class I), while RNA polymerase II transcribes protein-coding genes (class II) and RNA polymerase III transcribes 5S and tRNA genes (class III).

RNA polymerase II has been purified to near homogeneity from more that 20 organisms, from plant and animal tissues, cultured cells and lower eukaryotes (reviewed by Sawadago and Sentenac, 1990). Such studies have demonstrated that RNA polymerase II is a multi-subunit enzyme of molecular mass 500-600 KDa consisting of two large polypeptides (M_r > 140 KDa) and a number of smaller polypeptides ranging in size from 10 KDa to 40 KDa (reviewed by Sawadago and Sentenac, 1990). The genes encoding RNA polymerase II have been cloned from yeast cells (Ingles *et al.*, 1984), *Drosophila* (Ingles *et al.*, 1983) and mouse and human cells (Cho *et al.*, 1985).

The largest subunit of RNA polymerase II has an unusual COOH-terminal domain (CTD) composed of tandemly repeated copies of a serine-rich heptapeptide with the consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser (reviewed by Young, 1991). This CTD consensus sequence occurs 26-27 times in yeast RNA polymerase (Allison *et*

al., 1985; Nonet *et al.*, 1987) and 52 times in the enzyme from mammalian cells (Corden *et al.*, 1985; Allison et al., 1988). RNA polymerase II comes in two forms, IIA and IIO, which differ only in the extent with which their CTDs are phosphorylated. The CTD can be phosphorylated by a DNA-dependent protein kinase (DNA-PK) at multiple sites *in vivo* (Cadena and Dahmus, 1987). The IIA form is the non-phosphorylated state while the IIO form is phosphorylated. It is thought that the kinase which specifically phosphorylates the CTD may provide the signal for RNA polymerase II to initiate mRNA synthesis once it is assembled into a stable complex at the promoter (reviewed by Peterson and Tjian, 1992).

Genetic and biochemical experiments have demonstrated the functional importance of the CTD of RNA polymerase II in transcriptional initiation. Deletion mutants that remove most or all of the CTD are lethal to yeast, *Drosophila* and mouse cells (Nonet *et al.*, 1987; Allison *et al.*, 1988; Bartolomei *et al.*, 1988; Zehring *et al.*, 1988). Additionally, monoclonal antibodies specific for the CTD have been shown to inhibit transcription initiation from a variety of promoters *in vitro* (Dahmus and Kedinger, 1983). Other such monoclonal antibodies inhibit the formation of stable initiation-competent complexes and transcription initiation from the adenovirus major late (AdML) and mouse dihydrofolate reductase (DHFR) promoters *in vit*ro (Thompson *et al.*, 1989; Conaway *et al.*, 1992a).

The CTD of RNA polymerase II has been shown to interact with components of the general transcription factor TFIID (Conaway *et al.*, 1992a; Usheva *et al.*, 1992). As certain activators of transcription are thought to communicate with the basal transcriptional apparatus by interactions with TFIID (Section 1.2.3.2.) (Pugh and Tjian, 1990; Smale *et al.*, 1990; Dynlacht et al., 1991; White *et al.*, 1991; Pugh and Tjian, 1992) it has been suggested that trans-acting DNA binding proteins may regulate RNA polymerase II transcription through direct or indirect interactions with the CTD.

1.2.2.2. The general transcription factors.

The development of soluble extracts capable of reproducing *in vitro* the accurate transcription of class II promoters has been a major breakthrough for investigating both the general and regulatory mechanisms of transcription initiation by RNA polymerase II.

Studies of initiation in such extracts has resulted in the purification of many of the general transcription factors and the cloning of their genes. Accurate RNA polymerase II transcription can be reconstituted using highly purified components of the general transcription factors (TFIIA, -B, -D, -E, -F, -H and -J) which will be discussed in the following subsections.

1.2.2.2.a. TFIID and TBP.

TFIID, the pivotal factor in the basal transcription apparatus, is a multiprotein complex which binds to the TATA box by way of a critical subunit, the TATA-box binding protein (TBP).

TBP.

The first committed step in the formation of a complete preinitiation complex at the core promoter is the binding of TBP to a sequence element centred around the TATA box (reviewed by Sawadogo and Sentenac, 1990).

As the mammalian TBP is highly labile and difficult to purify (Nakajima *et al.*, 1988), most investigations initially centred around the yeast TBP protein. Purification of yeast TBP revealed that it is composed of a single 27 KDa polypeptide which is capable of binding specifically to the TATA region of many promoters and of substituting for human TBP in basal transcription (Buratowski *et al.*, 1988; Cavillini *et al.*, 1989; Hahn *et al.*, 1989a; Horikoshi *et al.*, 1989; Schmidt *et al.*, 1989a). The subsequent isolation of cDNA clones encoding the yeast TBP (Cavallini *et al.*, 1989; Eisenmann *et al.*, 1989; Hahn *et al.*, 1989a; Horikoshi *et al.*, 1989; Schmidt *et al.*, 1989a) led quickly to the isolation of cDNA clones encoding homologous proteins of approximately 37 KDa from a wide variety of species, including *Drosophila* (Hoey *et al.*, 1990; Muhich *et al.*, 1990), mouse (Tamura *et al.*, 1991) and humans (Hoffman *et al.*, 1990; Kao *et al.*, 1990; Peterson *et al.*, 1990).

Analysis of the deduced amino acid sequence of the different TBPs reveals that they are composed of a bi-partite structure, comprising a COOH-terminal domain of 180 amino acids, which shows a high degree of conservation between different organisms, and a non-conserved NH_2-terminal domain, varying in size in a species dependent manner (Greenblatt, 1991a; Pugh and Tjian, 1992). A series of studies have demonstrated that the COOH-terminal domain of TBP appears to constitute the functional core of the molecule, being required for a basal level of activity. Limited proteolysis to remove the NH_2-terminal domain of yeast TBP results in a COOH-terminal domain protein which retains the ability to bind effectively to minimal promoter elements and to activate basal levels of transcription (Lieberman *et al.*, 1991). Additionally, TBP mutants lacking the NH_2-terminal region have been shown to be fully functional in promoter binding and basal transcription (Hoey *et al.*, 1990; Peterson *et al.*, 1990; Kelleher *et al.*, 1992). Genetic evidence is consistent with the conserved COOH-terminal domain constituting the core of the molecule, as it has been shown for yeast TBP that the COOH-terminal domain contains all the functions essential for normal yeast cell growth and for response to acidic transcriptional activators *in vivo* (Cormack *et al.*, 1991; Gill and Tjian, 1991; Poon *et al.*, 1991; Reddy and Hahn, 1991; Zhou *et al.*, 1991).

The COOH-terminal domain is rich in basic amino acids and contains two 60-amino-acid direct repeats interrupted by a lysine-rich spacer region of 40 amino acids (Cavallini *et al.*, 1989; Hoejimakers, 1990; Nagai, 1990). Overlapping the more COOH-terminal direct repeat is a region similar in sequence to prokaryotic sigma factors (Horikoshi *et al.*, 1989; Hoffmann *et al.*, 1990; Jaehning, 1991). Sigma, a loosely associated component of the bacterial RNA polymerase, is required for accurate initiation of transcription and contains diverse activities such as RNA polymerase binding, promoter recognition and DNA melting (reviewed by Helmann and Chamberlin, 1988; Lonetto *et al.*, 1992).

The direct repeat sequences within the COOH-terminal domain of TBP have been shown to be required for promoter binding (Reddy and Hahn, 1991; Yamamota *et al.*, 1992) yet it appears that TBP cannot be divided into distinct DNA-binding and transcriptional activation domains. Studies have shown that deletion of amino acids anywhere within the entire COOH-terminal domain abolishes both promoter binding and transcriptional activation (Horikoshi *et al.*, 1990).

DNase 1 and chemical footprinting analysis have demonstrated that TBP binds to a small promoter region centred around the TATA box (Hahn *et al.*, 1989a, 1989b; Horikoshi *et al.*, 1989; Kao *et al.*, 1990) and unlike sequence-specific DNA binding proteins, TBP binds to and dissociates from TATA elements very slowly (Schmidt *et al.*, 1989b; Hoopes *et al.*, 1992). Although TBP is transcriptionally most active when bound to the consensus TATA element, TATAAA (Wobble and Struhl, 1990), it has been shown to bind to nearly all permutations of the canonical TATA sequence as well as a variety of nonconsensus TATA elements (Hahn *et al.*, 1989b; Singer *et al.*, 1990). In fact it has been shown that human TBP will even recognise and bind to DNA sequence elements located 30 bp upstream of several promoters, such as the simian virus 40 (SV40) major late promoter, which lack definable TATA elements, albeit with weaker affinities than it binds to consensus TATA box sequences (Wiley *et al.*, 1992).

TFIID.

Efforts to purify TFIID from higher eukaryotes have been hampered by its low abundance in cells and by its tendency to suffer degradation or dissociation into smaller transcriptionally active species during purification (Nakajima *et al.*, 1988). TFIID has subsequently been shown to be a high molecular mass, multi-subunit complex composed of TBP and additional polypeptides, termed TBP-associated factors (TAFs) (Pugh and Tjian, 1992). It has been proposed that the TAFs associated with TBP participate in promoter recognition and are required for transcriptional activation by upstream activator proteins (Dynlacht *et al.*, 1991; Greenblatt, 1991a; Tanese *et al.*, 1991; Pugh and Tjian, 1992).

It has been shown that TBP is essential for transcription from all three classes of eukaryotic genes and that its associations with TAFs appear to specify which polymerase can initiate transcription from a given gene (reviewed by Sharp, 1992). Thus, several forms of TFIID possessing molecular masses of approximately 700 KDa (Conaway *et al.*, 1990a; Zhou *et al.*, 1992), 300 KDa (Timmers and Sharp, 1991; Lobo *et al.*, 1992) and 230 KDa (Comai *et al.*, 1992) have been identified in extracts of human cells. The 700 KDa form appears to be required for transcription by RNA polymerase II from class II promoters, the 300 KDa form is termed B-TFIID and appears to be required for transcription from class III promoters lacking a TATA box (Lobo *et al.*, 1992) and the 230 KDa form appears to be involved in transcription by RNA polymerase I from class I promoters, as it was found to co-purify with the RNA polymerase I initiation factor SL1 (Comai *et al.*, 1992).

Using antibodies directed against TBP, attempts have been made to identify the factors associated with the TFIID complex by immunoaffinity purification procedures. Multiple TAFs which associate tightly with TBP have subsequently been identified. These consist of six prominent polypeptides in *Drosophila* cells, ranging in molecular mass from 31 to 150 KDa (Dynlacht *et al.*, 1991) and approximately ten polypeptides (10-250 KDa) in mammalian cell extracts (Pugh and Tjian, 1991; Tanese *et al.*, 1991; Zhou *et al.*, 1992). Zhou *et al.* (1992) reported the purification of the TFIID complex from HeLa cells. It was demonstrated that TFIID appears as a single purified multisubunit complex of about 700 KDa, termed Holo-TFIID, composed of TFIID and several TAFs ranging in size from 28 to 250 KDa. Holo-TFIID appeared to represent the native TFIID as it was capable of supporting *in vitro* transcription from an RNA polymerase II promoter (Zhou *et al.*, 1992).

The 250 KDa subunit of the TFIID complex has subsequently been identifed as the cell-cycle regulatory protein CCG1 (Hisatake *et al.* 1993a; Ruppert *et al.*, 1993). CCG1 is a nuclear DNA-binding protein which is important for cell cycle progression (Sekiguchi *et al.*, 1988, 1991). CCG1 contains regions of sequence displaying homologies with known DNA-binding and transcriptional activation domains, including an HMG box (Section 1.2.3.1.1.j.) and a domain rich in proline residues (Section 3.1.2.3.). On the basis of these homologies, it was suggested that CCG1 might be a sequence-specific transcription factor involved in the regulation of genes controlling progression through the cell cycle (Sekiguchi *et al.*, 1991). Based on its ability to interact with TBP it has now been suggested that CCG1, in addition to serving a major cell cycle function, forms part of the TFIID complex and may play a role in both recruiting other TAFs to the complex and in receiving signals from various transcriptional regulators (Hisatake *et al.* 1993a; Ruppert *et al.*, 1993).

Studies using *in vitro* transcription from various promoters have shown that whereas TBP requires only a short sequence element centred around the TATA motif to direct transcription, TFIID appears to recognise and interact with a large promoter element. Both the initiator element (Section 1.2.2.4.) and the TATA box motif were shown to be required for transcription from the AdML promoter using rat liver extract, however when yeast TBP was used only the TATA box was required for efficient transcription (Conaway *et al.*, 1990b, 1991). Similarly, transcription driven from a synthetic promoter was shown to require both the TATA box and the initiator elements when human TFIID was used, yet only sequences in the vicinity of the TATA box were required for transcription directed by human and yeast TBP (Nakatani *et al.*, 1990; Smale *et al.*, 1990). *In vitro* binding data using DNase 1 footprinting analysis of the AdML and human histone 4 promoters is also consistent with TFIID interacting with a larger region of core promoters than does TBP (Sawadogo and Roeder, 1985b; Nakajima *et al.*, 1988; Van Dyke and Sawadogo, 1990).

From the observation that both TBP and TFIID can interact with the TATA box motif, it seems likely that the ability of TFIID to recognise a larger region of minimal promoters than TBP may be accomplished through the co-operative interactions of TBP with the TATA box and of the TAFs with initiator elements and other parts of the minimal promoter. TFIID has been shown to direct transcription from the core regions of promoters lacking definable TATA boxes (TATA-less) and it is predicted that the binding of TFIID to these TATA-less promoters is independent of TBP and is achieved through the actions of the TAFs (Pugh and Tjian, 1990, 1991, 1992). The role of the TBP-subunit of TFIID in such TATA-less promoters is unclear but it has been suggested that human TBP will bind weakly, but specifically, to sequences located upstream of the transcriptional start site of such promoters (Wiley *et al.*, 1992). Therefore, it may have a qualitatively similar function to that at TATA-box containing promoters in correctly positioning RNA polymerase II and the transcription apparatus at minimal promoter elements.

Thus, it appears that TBP, in association with accessory factors, forms the TFIID complex and it is this factor which binds stably to the minimal promoter regions of TATA-box containing promoters to form the first stage of transcriptional initiation by RNA polymerase II.

1.2.2.2.b. TFIIA.

TFIIA was originally identified as a HeLa cell-derived factor possessing transcriptional stimulatory properties (Matsui *et al.*, 1980; Samuels and Sharp, 1986). Reports describing the purification of TFIIA from mammalian sources revealed variable subunit compositions, consisting of a single 43 KDa protein (Egly *et al.*, 1984), a 38

KDa protein (Usuda *et al.*, 1991), two subunits of 12 and 19 KDa (Samuels and Sharp, 1986), or a heterotrimer containing of subunits of 12, 19 and 35 KDa (Cortes *et al.*, 1992; Coulombe *et al.*, 1992). TFIIA purified from yeast exists as subunits of 43 and 12 KDa (Hahn *et al.*, 1989c; Ranish and Hahn, 1991). The two yeast genes encoding these subunits have subsequently been cloned and the recombinant proteins are transcriptionally active in yeast cells and will function interchangeably with the mammalian factors in basal transcription (Ranish *et al.*, 1992). Recently, a cDNA clone encoding the largest subunit (α) of human TFIIA has been cloned (Ma *et al.*, 1993). A separate study reported the isolation of a human cDNA clone, hTFIIA/α, which encodes a 55 KDa protein (DeJong and Roeder, 1993). This single cDNA was shown to contain the sequences of both the 19 KDa and the 35 KDa subunits and therefore encodes both of these subunits (DeJong and Roeder, 1993).

The role of TFIIA in directing accurate transcription by RNA polymerase II has remained elusive. It has been shown that TFIIA promotes stable binding of TFIID to the core promoter (Davison *et al.*, 1983; Reinberg *et al.*, 1987; Buratowski *et al.*, 1989) and that it will bind stably to both TBP (Ranish and Hahn, 1991; Usuda *et al.*, 1991; Cortes *et al.*, 1992; Ma *et al.*, 1993) and to TFIID (Ma *et al.*, 1993). However, reconstituted transcription systems have failed to show a consistent requirement for TFIIA. The *in vitro* requirement of TFIIA for the synthesis of specific transcripts ranges from no requirement (Sawadogo and Roeder, 1985a; Van Dyke *et al.*, 1988; Conaway *et al.*, 1990a; Sayre *et al.*, 1992), to a stimulatory effect (Egly *et al.*, 1984; Cortes *et al.*, 1992) or even an absolute requirement (Samuels and Sharp, 1986; Reinberg *et al.*, 1987). It appears that the requirement of this factor for basal transcription may depend on both its source and on the degree of its purification. No requirement for TFIIA is observed in highly purified systems from yeast (Sayre *et al.*, 1992) and rat (Conaway *et al.*, 1990a), yet it was found to be required in a reconstituted transcription system from human cells when a partially purified preparation of TFIID was used. However when this was replaced by purified TBP, TFIIA was no longer required (Cortes *et al.*, 1992). Additionally, mutations to TBP which abolish its ability to interact with TFIIA have no effect on TBPs ability to support *in vitro* transcription from the AdML promoter (Buratowski and Zhou, 1992; Yamamoto *et al.*, 1992).

Thus, at present thereareconflicting data concerning the role of TFIIA as a basal transcription factor.

1.2.2.2.c. TFIIB.

A great deal of interest has recently been centred around the transcription factor TFIIB. The reason for this interest stems from the fact that TFIIB is thought to be one

of the main factors involved in the transcription of DNA by RNA polymerase II and it is also thought to be involved as a target of trans-acting activator proteins.

TFIIB was initially purified to homogeneity from rat liver (Conaway *et al.*, 1987) and subsequently from human cells where it exists as a single 35 KDa polypeptide (Ha *et al.*, 1991; Malik *et al.*, 1991; Moncollin *et al.*, 1992). cDNAs encoding the human (Ha *et al.*, 1991; Malik *et al.*, 1991), *Drosophila* (Wampler and Kadonaga, 1992) and rat (Tsuboi *et al.*, 1992) factors have since been isolated.

The sequence of TFIIB from humans and other organisms has several notable features. At the NH_2-terminus there is a conserved cysteine-rich sequence postulated to form a Zn^{2+}-binding region. The COOH-terminus mostly comprises two imperfect repeats, each of about 75 amino acids, with a cluster of basic residues at the end of the most NH_2-terminal repeat spaced in such a manner that they could potentially form an amphipathic α-helix.

It has been shown that TFIIB will form a complex with TFIID through protein-protein interactions (Lin and Green, 1991). TFIIB is also thought to bind a protein complex consisting of RNA polymerase II and TFIIF thereby facilitating their binding to the core promoter region (Tschochner *et al.*, 1992). The fact that TFIIB appears to act as a bridging factor, using protein-protein contacts to bind both TBP and the RNA polymerase II-TFIIF complex, has led to the suggestion that it may consist of two functionally distinct domains. Deletion analysis has subsequently shown that the COOH-terminal region, containing the direct repeats and associated basic regions, is necessary and sufficient for interaction of TFIIB with TBP but is inactive in transcription (Barberis *et al.*, 1993; Buratowski and Zhou, 1993; Hisatake *et al.*, 1993b). By contrast, the NH_2-terminal region, containing the putative Zn^{2+}-binding domain, was found to be dispensable for binding to the TBP-DNA complex but was shown to be required for transcriptional activity. It was shown that the NH_2-terminal region of TFIIB is probably involved in the incorporation of RNA polymerase II-TFIIF into the preinitiation complex through protein-protein interactions with RNA polymerase II, TFIIF, or both (Barberis *et al.*, 1993; Buratowski and Zhou, 1993; Hisatake *et al.*, 1993b).

1.2.2.2.d. TFIIF.

TFIIF (also termed RAP 30/74 and β/γ) appears to exist as a heterodimer composed of subunits of relative molecular mass 30 and 58 KDa (termed RAP 30 or β and RAP 74 or γ respectively). The subunits constituting TFIIF were first identified on the basis of their ability to bind immobilized RNA polymerase II (reviewed by Greenblatt, 1991b) and subsequently were purified from rat liver (Conaway and Conaway, 1989a) and human cells (Flores *et al.*, 1990; Kitajima *et al.*, 1990). The

genes encoding both subunits have now been cloned (Sopta *et al.*, 1989; Aso *et al.*, 1992; Finkelstein *et al.*, 1992), and both have been shown to be required for transcriptional activity from a number of cellular genes (Conaway and Conaway, 1989a; Flores *et al.*, 1990; Kitajima *et al.*, 1990; Aso *et al.*, 1992; Finkelstein *et al.*, 1992). TFIIF does not appear to bind DNA and contrary to original reports (Sopta *et al.*, 1989), it does not appear to posses an associated DNA helicase activity (Flores *et al.*, 1990; Kitajima *et al.*, 1990; Finkelstein *et al.*, 1992)

It has been shown that TFIIF can associate with RNA polymerase II in solution (Sopta *et al.*, 1989; Flores *et al.*, 1990; Kitajima *et al.*, 1990) and the smaller subunit, RAP 30, has been shown to bind directly to RNA polymerase II (McCraken and Greenblatt, 1991; Killeen and Greenblatt, 1992). Sequence analysis of RAP 30 has indicated homology with bacterial sigma factor in the RNA polymerase-binding region, and both factors appear functionally homologous (Sopta *et al.*, 1989; McCraken and Greenblatt, 1991; Killeen and Greenblatt, 1992). Whether the RAP 74 subunit also binds to RNA polymerase II is unclear. Recombinant RAP 30 has been shown to bind immobilised RNA polymerase II in the absence of RAP 74 suggesting that TFIIF binds solely through the RAP 30 subunit (Killeen and Greenblatt, 1992). However, other studies have provided evidence which suggests that RAP 74 acts to stabilise the interaction of RAP 30 with RNA polymerase II (Garrett *et al.*, 1992).

Another region of RAP 30, which also shares sequence homology with bacterial sigma, has shown to be required for transcriptional activity (Garrett *et al.*, 1992). This particular region in bacterial sigma has been established as playing a role in the recognition and binding to TATA-like sequences of prokaryotic promoters (Lonetto *et al.*, 1992). Although purified TFIIF does not bind DNA, this region may participate in promoter recognition or DNA binding in the context of the preinitiation complex.

TFIIF has been shown to prevent non-specific binding of RNA polymerase II to free DNA and has also been shown to dissociate non-productive pre-formed complexes between RNA polymerase II and DNA (Killeen and Greenblatt, 1992). Again bacterial sigma plays a similar role to this in preventing the non-specific binding of *E.coli* RNA polymerase to non-promoter sites on DNA. How TFIIF acts to suppress such non-specific binding is unclear but it may be a property of the RAP 30 subunit since it can be achieved with recombinant RAP 30 produced in *E.coli* (Killeen and Greenblatt, 1992). However, RAP 30 alone is incapable of dissociating non-productive RNA polymerase II-DNA complexes (Killeen and Greenblatt, 1992), suggesting that both RAP 30 and RAP 74 may play an important role in controlling the binding of RNA polymerase II to DNA.

1.2.2.2.e. TFIIH.

Purified rat TFIIH (or δ) appears to have a native molecular mass of 400 KDa and consists of eight polypeptides, ranging in molecular weights from 35 to 94 KDa (Conaway and Conaway, 1989b; Conaway *et al.*, 1992b). Human TFIIF (or BTF2) appears to have a native molecular mass of 200 KDa and also appears to consist of eight polypeptides, ranging from 32 to 90 KDa (Gerard *et al.*, 1991, Fischer *et al.*, 1992).

TFIIH appears to posses both kinase and transcriptional activities (Lu *et al.*, 1992). As already mentioned, the CTD of the largest subunit of RNA polymerase II is a substrate for phosphorylation (Section 1.2.2.1.). It has been shown that the non-phosphorylated form of RNA polymerase II (IIA) is recruited to the preinitiation complex (Lu *et al.*, 1991) and that the CTD is subsequently phosphorylated to produce the transcriptionally active IIO form (Cadena and Dahmus, 1987; Payne *et al.*, 1989). Lu *et al.* (1992) demonstrated that TFIIH contains a kinase activity capable of phosphorylating the CTD of RNA polymerase II. This activity was demonstrated in absence of any other factors, although factors which promote the association of RNA polymerase II into the preinitiation complex greatly stimulated the TFIIH-kinase activity (Lu *et al.*, 1992).

It appears that transcription and kinase activity are both contained within the multisubunit TFIIH. TFIIH separates into two fractions on phosphocellulose and both fractions have been shown to posses transcriptional and kinase activities (Lu *et al.*, 1992). Additionally, monoclonal antibodies directed against the 62 KDa subunit of TFIIH act to inhibit both transcriptional and kinase activities (Lu *et al.*, 1992).

Therefore, TFIIH clearly has a crucial role to play in assembly of the preinitiation complex. It has been shown to enter the complex (Conaway *et al.*, 1992b), associate stably with RNA polymerase II in solution (Gerard *et al.*, 1991) and contains both a kinase activity and transcriptional activity.

1.2.2.2.f. TFIIE.

Human and rat TFIIE are thought to exist as heterotetramers composed of subunits of 34 and 56 KDa ($\alpha_2\beta_2$) (Ohkuma *et al.*, 1990; Conaway *et al.*, 1991; Inostroza *et al.*, 1991). The cDNAs encoding both subunits have been isolated (Ohkuma *et al.*, 1991; Peterson *et al.*, 1991; Sumimoto *et al.*, 1991) and initial reports suggested that the 56 KDa subunit possesses the transcriptional activity with the 34 KDa subunit having a stimulatory role (Ohkuma et al., 1990; Inostroza *et al.*, 1991). However, it now appears that both subunits are required for optimal reconstituted basal-level transcription (Conaway *et al.*, 1991; Ohkuma *et al.*, 1991; Peterson *et al.*,

1991; Sumimoto *et al.*, 1991) and both are also sufficient to reconstitute Sp1-activated transcription (Peterson *et al.*, 1991).

TFIIE does not appear to bind DNA on its own (Ohkuma et al., 1990; Conaway *et al.*, 1991; Inostroza *et al.*, 1991) and although the 56 KDa subunit possesses a sequence with similarity to a highly conserved sequence found in protein kinases, no DNA-dependent ATPase activity has been reported for purified or recombinant TFIIE (Buratowski *et al.*, 1989; Inostroza *et al.*, 1991; Peterson *et al.*, 1991; Cortes *et al.*, 1992; Flores *et al.*, 1992). TFIIE has, however, been shown to associate stably with the preinitiation complex after RNA polymerase II and TFIIF (Inostroza *et al.*, 1991; Peterson *et al.*, 1991) and reports suggest that it will associate stably with RNA polymerase II and TFIIF in solution (Flores *et al.*, 1989; Buratowski *et al.*, 1991).

1.2.2.2.g. TFIIG and TFIIJ.

TFIIJ was found to co-purify with TFIIA through several chromatographic steps (Cortes *et al.*, 1992) and appears to be required for transcription directed by yeast or human TBP (Cortes *et al.*, 1992).

TFIIG has been shown to be chromatographically distinct from the human factors TFIIA, TFIIB, TFIIF, and TFIID (Sumimoto *et al.*, 1990) and may in fact composed of a mixture of TFIIH and TFIIJ (Flores *et al.*, 1992).

As neither factor has been purified to homogeneity as yet, little more is known about the characteristics of each or their potential roles in the initiation of transcription.

1.2.2.3. Mechanisms of transcriptional initiation by RNA polymerase II.

Various approaches have been used to identify and characterise the intermediate steps in the complex reaction pathway leading to specific transcription initiation by RNA polymerase II. These approaches include footprinting analysis (Van Dyke *et al.*, 1988), native gel electrophoresis (Buratowski *et al.*, 1989) and studies carried out in reconstituted transcription systems. Based on these studies a working model has evolved for the assembly of the functional preinitiation complex involving a multistage process and the action of at least five initiation factors and an ATP cofactor. A diagrammatic representation of this working model is shown in Figure 2 and is described below.

Step 1. TBP binds stably to the core promoter around the TATA box region to form an initial complex (reviewed by Sawadogo and Sentenac, 1990).

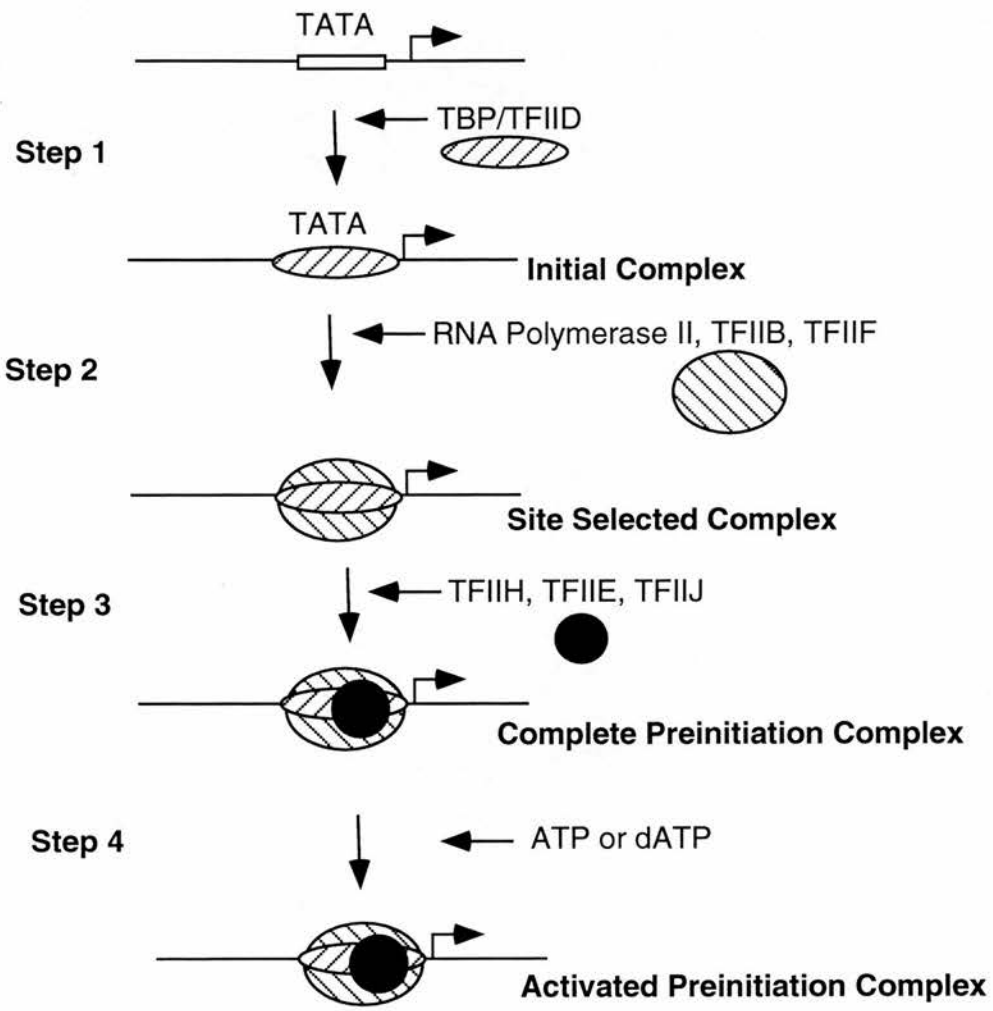
Step 2. The initial complex serves as the recognition site for RNA polymerase II (Fire *et al.*, 1984; Reinberg *et al.*, 1987) which, assisted by at least two additional initiation factors, TFIIB (Buratowski *et al.*, 1989) and TFIIF (Flores *et al.*, 1991), binds productively to the initial complex to form the site-selected complex.

Figure 2. A model for assembly of the functional preinitiation complex.

Initiation of transcription by RNA polymerase II from the core regions of TATA box-containing promoters is a multistage process requiring the action of the general transcription factors, TBP/TFIID, TFIIB, TFIIE, TFIIF, TFIIH, TFIIJ and an ATP cofactor.

The order in which the components of the functional preinitiation complex are thought to be recruited onto the template is shown as steps 1 to 4. These factors and their assembly onto the template are discussed in more detail in Sections 1.2.2.2. and 1.2.2.3.

The TATA box is indicated and the transcriptional start site (+1) is marked by an arrow. TBP denotes TATA-box binding protein and is part of the larger TFIID protein complex (see Section 1.2.2.2.a.).


TATA
TBP/TFIID
Step 1
TATA
Initial Complex
RNA Polymerase II, TFIIB, TFIIF
Step 2
Site Selected Complex
TFIIH, TFIIE, TFIIJ
Step 3
Complete Preinitiation Complex
ATP or dATP
Step 4
Activated Preinitiation Complex

Step 3. RNA polymerase II then, in association with the initiation factors TFIIH (Cortes *et al.*, 1992; Flores *et al.*, 1992), TFIIE (Cortes *et al.*, 1992; Flores *et al.*, 1992) and TFIIJ (Cortes *et al.*, 1992), appears to extend its interactions downstream to form the complete, but inactive, preinitiation complex (Van Dyke *et al.*, 1988; Buratowski *et al.*, 1989; Flores *et al.*, 1991).

Step 4. In an ATP (dATP)-dependent step this preinitiation complex is converted to an activated complex, capable of initiating transcription rapidly when provided with the remaining ribonucleoside triphosphates. ATP has been proposed to serve two functions in this step. It is proposed to promote conversion of the complete preinitiation complex from a 'closed' to an 'open' conformation by providing the energy to drive unwinding of a short stretch of promoter DNA surrounding the transcriptional start site (Wang *et al.*, 1992). ATP is also proposed to act as a substrate for protein kinases that phosphorylate the CTD of the largest subunit of RNA polymerase II (Section 1.2.2.1.).

Thus, it appears that the initiation of transcription by RNA polymerase II requires the action of a number of initiation factors. These are thought to act in concert to promote selective binding of RNA polymerase II to the core promoter and subsequently to promote conversion of the site-selected complex to the complete complex. Exactly how assembly of the complete preinitiation complex is promoted by the initiation factors is not fully understood, but it is thought to be a complex process involving multiple steps and factors. Evidence suggests that at least some of the initiation factors, individually or in combination, are capable of interacting with RNA polymerase II in the absence of DNA (Flores *et al.*, 1989; Buratowski et al., 1991), therefore it is possible that RNA polymerase II assembles with all or a subset of its initiation factors prior to entering the initiation complex.

1.2.2.4. Initiator elements.

Initiator (Inr) elements are found in the vicinity of the transcription initiation site of several eukaryotic genes, both containing TATA boxes, such as the histone H2A promoter (Grosschedl and Birnstiel, 1980), the Ad-ML promoter (Smale *et al.*, 1990; Roy *et al.*, 1991), the adeno-associated virus type 2 promoter (Seto *et al.*, 1991) and the HIV-1 promoter (Du *et al.*, 1993) and also in genes lacking well-defined TATA boxes, such as the lymphocyte-specific terminal deoxynucleotidyltransferase (TdT) promoter (Corden *et al.*, 1980) and the SV40 late promoter (Ayer and Dynan, 1988).

Some promoters contain multiple Inr elements, such as the HIV-1 and Ad-ML promoters (Du *et al.*, 1993) which contain two such elements located 3' of the transcriptional initiation site, both contributing to promoter strength *in vitro* and *in vivo*.

The Inr sequence has been found to be both necessary and sufficient to direct basal transcription from several promoters, including the TATA-less TdT promoter

(Smale and Baltimore, 1989) and the TATA-box containing Ad-ML (Roy *et al.*, 1991; Du *et al.*, 1993) and adeno-associated virus type 2 promoters (Seto *et al.*, 1991).

Multiple types of Inr elements are thought to exist as not all elements are homologous in sequence (Ayer and Dynan, 1988; Smale and Baltimore, 1989) and electrophoretic mobility shift assays, using different Inr elements as oligonucleotide probes, have shown that different complexes are formed which display distinct mobilities and antigenic cross-reactivity (Seto *et al.*, 1991). A weak consensus sequence for Inr elements has however been defined (Corden *et al.*, 1980), as shown below:

5'-PyPyCAPyPyPyPyPy-3'

Several factors have been shown to exist which specifically recognise Inr elements from a range of different promoters. These include upstream stimulatory factor (USF) and TFII-I (Roy *et al.*, 1991; Du et al., 1993) which recognise the AdML, HIV-1 and TdT Inr elements, a factor termed YY1 initially identified as binding to the adeno-associated virus type 2 Inr element (termed P5+1) (Seto *et al.*, 1991; Shi *et al.*, 1991) and the Inr-binding protein of SV40, termed IBP-s (Wiley *et al.*, 1993). A recent study investigated the properties of multiple types of Inr elements and demonstrated that they all appear to be recognised by the same universal binding protein (Javahery *et al.*, 1994). It was proposed from this study that the different proteins reported to bind Inrs such as USF, TFII-I and YY1 may augment promoter strength and act in conjunction with a universal protein (Javahery *et al.*, 1994).

Different Inr elements have been shown to both activate or repress transcription from different promoters. For example, IBP-s has been shown to act as a cellular repressor of transcription (Wiley *et al.*, 1993) and it has been shown that certain other identified Inr elements, such as those from the HIV-1 (Kato *et al.*, 1991) and adeno-associated virus type 2 promoters (Seto *et al.*, 1991; Shi *et al.*, 1991), can act to both activate or repress transcription depending on their context within the promoter. It was proposed from these findings that the factors which bind to Inr elements may interact in different ways with TFIID, RNA polymerase II and other components of the general transcriptional machinery depending on the exact location of each element. Factors which bind around the transcriptional start site are postulated to interact with the general transcription factors to enhance transcription and factors which bind to TATA box elements are postulated to block the binding of TFIID, thereby repressing transcription (Kato *et al.*, 1991).

1.2.3. Upstream regulatory sequences.

Upstream activating sequences (UASs) and enhancer elements act to regulate the frequency of transcriptional initiation by RNA polymerase II at the minimal promoter.

UASs are typically short (8-12 bp) gene-specific motifs which act as binding sites for transcriptional activators (upstream element binding proteins) (reviewed by Guarente, 1987; Johnson and McKnight, 1989; Struhl, 1989). Such motifs include CCAAT box elements (Christy *et al.*, 1989; Friedman *et al.*, 1989; Crossley and Brownlee, 1990) and Sp1 binding sites (Jones and Tjian, 1985). UASs can act to stimulate transcription either constitutively or in response to specific stimuli such as serum stimulation (Triesman, 1985), heat shock (Bienz and Pelman, 1986) or steroids (Renkawitz *et al.*, 1984). They can operate in either orientation, but show a decrease in activity with increasing distance from the TATA box (McKnight and Kingsbury, 1982).

Enhancers are defined by their ability to act over considerable distances, both upstream and downstream of the transcriptional start site, and by their independence of orientation in relation to the gene. They can be longer than UASs, sometimes comprising hundreds of bases (Edlund *et al.*, 1985), and can contain binding sites for many different transcription factors (Serfling *et al.*, 1985).

1.2.3.1. Transcriptional activators.

One of the largest and most diverse classes of DNA-binding proteins are the transcription factors dedicated to binding specific DNA sites (or sets of sites) and to regulating the transcriptional machinery.

Such transcription factors are modular in nature being comprised of specific domains. They posses a domain which allows them to recognise and interact with specific DNA sequences, a separate and independent region required for the regulation of transcription initiation and, in some cases, a domain that favours protein-protein interactions (Ptashne, 1988). Consensus sequences can be deduced for sets of transcription factors which bind to similar DNA sequence elements.

1.2.3.1.1. DNA-binding domains.

The structural characteristics have been reported for more than ten types of protein-DNA complexes and for more than twenty types of DNA-binding proteins. From structural studies and sequence comparisons, it has emerged that many DNA-binding proteins can be grouped into classes that use related structural motifs for the specific recognition of DNA sequences.

This section will give an overview of the major families of DNA-binding proteins characterised to date which include the helix-turn-helix proteins, the winged-helix proteins, the homeodomains and POU domains, zinc finger proteins, leucine zipper proteins, the helix-loop-helix proteins, helix-span-helix proteins and HMG-domain proteins.

1.2.3.1.1.a Helix-turn-helix.

The helix-turn-helix (HTH) was the first DNA-recognition motif discovered and was initially characterised for the λ Cro protein (Anderson *et al.*, 1981; Ohlendorf *et al.*, 1982), the *E. coli* CAP protein (McKay and Steitz, 1981) and the DNA-binding domain of λ repressor (Pabo and Lewis, 1982).

From a comparison of the sequence of these three proteins, a conserved recognition domain, termed the HTH motif, was described consisting of an α-helix, a turn, and a second α-helix, often called the recognition helix, with the helices crossing each other at an angle approaching 120° (Steitz *et al.*, 1982; Ohlendorf *et al.*, 1983).

Subsequent sequence comparisons have demonstrated that this motif is present in a large family of prokaryotic DNA-binding proteins (Matthews *et al.*, 1982; Sauer *et al.*, 1982; Weber *et al.*, 1982). From such comparisons it appears that the most highly conserved residues include a glycine in the turn and several hydrophobic residues. These residues are believed to stabilise the arrangement of the two helices in the HTH motif and to help the motif pack against the rest of the protein. Sequence comparisons have also shown that the HTH motif always occurs as part of a larger DNA-binding domain with regions outside the HTH DNA-binding motif, such as an extended NH_2-terminal arm of the λ repressor, often being important for DNA sequence recognition (Jordan and Pabo, 1988, Clarke *et al.*, 1991).

The mechanism by which prokaryotic HTH proteins specifically recognise their DNA target sites, termed operators, has been investigated by resolving the structures of several repressor-operator complexes, including the λ repressor (Jordan and Pabo, 1988), λ Cro (Brennan *et al.*, 1990) and *E. coli* CAP (Schulz *et al.*, 1991) complexes. From such cocrystal structures it has been shown that HTH proteins appear to bind DNA as dimers with the conserved HTH motif of each monomer contacting the DNA in each half of the operator site.

In addition to possessing DNA-binding domains, HTH proteins often contain domains involved in regulating activity. For example, the NH_2-terminal domain of CAP allows dimer formation and also binds cAMP, which acts to allosterically alter DNA binding (McKay and Steitz, 1981). Additionally, the COOH-terminal domains of λ434 and LexA repressors allow stable dimer formation and are involved in the process of induction (Sauer *et al.*, 1990).

Variations on the canonical structure of the prokaryotic HTH motif have been discovered. For example, HTH motifs are also found in the eukaryotic homeodomain proteins (Section c) *engrailed* (Kissinger *et al.*, 1990), *Antennapedia* (Otting *et al.*, 1990) and MATα2 (Wolberger *et al.*, 1991), in the non-classical homeodomain hepatocyte transcription factor LFB/HNF1 (Ceska *et al.*, 1993; Leiting *et al.*, 1993), in the c-myb proto-oncogene (Ogata *et al.*, 1992), in the POU-specific domain of Oct-1 (Assa-Munt *et al.*, 1993; Dekker *et al.*, 1993) (Section d) and in the hepatocyte nuclear factor 3 (HNF-3) fkh family (Clark et al., 1993) (Section b). These proteins together form the class of variant HTH proteins.

1.2.3.1.1.b. Winged helix.

Monomeric variant HTH proteins, such as hepatocyte nuclear factor 3γ (HNF-3γ), have been shown to posses a novel DNA-binding motif termed the winged-helix motif (Clark *et al.*, 1993). HNF-3γ belongs to the HNF-3 fkh family of proteins which includes HNF-3α, HNF-3β and the gene products of the Drosophila *fkh* and *sloppy paired* 1 and 2 (*slp*1 and *slp*2) (Clark *et al.*, 1993). The members of this family share a highly conserved 110 residue DNA-binding domain which takes up an α/β structure composed of three NH_2-terminal helices, a three stranded antiparallel β sheet and three loops connecting the helices and the strands (Clark *et al.*, 1993). The crystal structure of the DNA-binding domain of HNF-3γ complexed to its DNA target site reveals, in addition to a HTH-like motif, the presence of a novel winged-helix motif which contacts the distal parts of the target site and is composed of an α-helix flanked on both sides by a loop.

1.2.3.1.1.c. Homeodomain.

The homeodomain is a DNA-binding motif which is present in a large family of eukaryotic regulatory proteins. The conserved residues constituting the homeodomain were first recognised in proteins important for regulating *Drosophila* development, such as *Antennapedia* (*Antp*) (Qian *et al.*, 1989; Billeter *et al.*, 1990). It is now known that the homeodomain has a much broader role in the regulation of eukaryotic gene regulation as homeodomain containing proteins have been found in virtually all eukaryotes examined ranging from yeast to man (reviewed by Scott *et al.*, 1989).

The homeodomain forms a stable, folded structure and is able to bind DNA by itself (Sauer *et al.*, 1988; Qian *et al.*, 1989; Affolter *et al.*, 1990). From the crystal structures of the *Drosophila engrailed* and the yeast α2 homeodomain-DNA complexes (Kissinger *et al.*, 1990; Wolberger *et al.*, 1991) it has been shown that both complexes have similar structures consisting of an extended NH_2-terminal arm and three α-helices.

Proteins possessing homeodomains which vary in sequence from the typical homeodomain have been shown to exist. For example, HNF1 a homeodomain-containing transcription factor, essential for the liver-specific expression of many hepatic genes, contains a divergent homeodomain possessing extra sequences upstream of the third helix (Nicosia *et al.*, 1990).

As mentioned in Section a, homeodomain proteins also belong to the class of HTH variant proteins. Amino-acid sequence comparisons have shown that the second and third helices of the homeodomain are similar to the HTH motif of prokaryotic repressors, although the homeodomain helices are longer (Laughon and Scott, 1984, Shepherd *et al.*, 1984). Additionally, structural studies of the complexes (Kissinger *et al.*, 1990; Otting *et al.*, 1990; Wolberger *et al.*, 1991) have shown that the homeodomain uses the HTH motif in a novel way. A comparison of the λ repressor-operator and engrailed homeodomain-DNA complexes (Kissinger *et al.*, 1990) shows that the two HTH units dock against DNA in different ways with critical base contacts being made by different residues from different portions of the motifs.

Although homeodomains, in isolation, have been shown to possess the ability to fold correctly and to bind DNA with a similar specificity to that of the intact protein (Sauer *et al.*, 1988; Qian *et al.*, 1989; Affolter *et al.*, 1990), it seems likely that the precise DNA-binding specificity is modulated by other regions of the protein. Many homeodomain proteins contain distinct sequence motifs which flank the homeodomain and are conserved within specific subfamilies (reviewed by Scott *et al.*, 1989). For example, the POU-domain proteins (Rosenfeld, 1991) possess neighbouring sequences which play important roles in DNA recognition (Section d). Conserved sequences contained within other homeodomain proteins may also have important roles in DNA recognition, thus modulating their specificity and/or affinity for DNA binding.

Protein-protein interactions may also play an important role in modulating the interactions of homeodomain proteins with DNA. For example, the yeast homeodomain protein α2, in addition to forming homodimers, can also form heterodimers with the related homeodomain protein a1 and with the yeast transcription factor Mcm1 (Goutte and Johnson, 1988; Kelleher *et al.*, 1989). Each of the formed complexes has a different sequence specificity and may interact with DNA in different ways. The hepatocyte transcription factor HNFI also binds to its target DNA sequence as a heterodimer formed with the homologous homeodomain protein vHNF1 (Nicosia *et al.*, 1990; Rey-Campos *et al.*, 1991) and the Paired/Pax class of homeodomain proteins have been shown to bind cooperatively as both homo- and heterodimers to palindrominc DNA sequences (Wilson *et al.*, 1993). Thus it appears that heterodimerization between homeoproteins also occurs in higher eukaryotes, potentially expanding their regulatory spectrum.

1.2.3.1.1.d. POU-domain.

The highly conserved POU-domain was initially recognised through analysis of the primary sequence of three mammalian transcription factors, Pit-1, Oct-1 and Oct-2, and a *Caenorhabditis elegans* (*C. elegans*) developmental regulator, unc 86. cDNAs encoding the pituitary specific transcription factor Pit-1, the B-cell specific octamer binding protein Oct-2 and the ubiquitous octamer binding protein Oct-1 were identified and cloned (Bodner *et al.*, 1988; Clerc *et al.*, 1988; Ingraham *et al.*, 1988; Ko *et al.*, 1988; Muller-Immergluck *et al.*, 1988; Scheidereit *et al.*, 1988; Sturm *et al.*, 1988). These three mammalian proteins and unc 86, a regulator of cell fate in *C. elegans* encoded by the *unc-86* gene (Chalfie *et al.*, 1981; Finney *et al.*, 1988), were found to share an extensively conserved domain of approximately 150 amino acids, termed the POU domain (Pit, Oct, Unc) (Herr *et al.*, 1988). Subsequently, numerous additional POU-domain genes have been identified in mammals, *Drosophila* and *C. elegans* where they have been postulated to serve a number of different functions and show varying and overlapping tissue distributions (reviewed in Rosenfeld, 1991).

The POU domain contains two major regions of high homology, a POU-specific (POU_S) domain and a POU-homeodomain (POU_{HD}) (Herr *et al.*, 1988). The POU_S domain lies immediately NH_2-terminal to the homeodomain and consists of a 65 to 75 amino acid residue segment. It can be further subdivided into two distinct regions of particularly high homology, the POU_S-A region and the POU_S-B region. The solution structure of POU_S has been determined by NMR and consists of four α-helices connected by short loops (Dekker *et al.*, 1993). A linker corresponding to a poorly conserved region of about 20 amino acids separates the POU_S domain from the 60 amino acid-long POU_{HD}. The POU_{HD} is distantly related to the classical homoedomain encoded by the homeobox. It is predicted to contain three α-helices corresponding to those in the classical homeodomains, with the third helix often being referred to as the 'recognition' helix (Herr *et al.*, 1988). Mutational analysis has demonstrated that the three α-helices from the POU_{HD} carry out different functions than the corresponding regions of the classical homeodomain proteins (Elsholtz *et al.*, 1990; Ingraham *et al.*, 1990).

On the basis of the primary sequence throughout the POU-domain, including the linker region, the POU-domain proteins have been classified into five groups (POU-I to POU-V) (reviewed by Rosenfeld, 1991). The majority of POU-domain proteins subsequently isolated can be categorised into one of the five classes, however some of the POU-domain proteins appear to posses divergent POU-domain sequences and may constitute novel classes. For example, a recently isolated and characterised zebra fish POU-domain protein, which is ubiquitously expressed during

embryogenesis, has been shown to posses an unusually divergent POU-domain (Johansen *et al.*, 1993). A novel POU domain protein has also been reported, termed TCFβ_1 (T-cell Receptor Factor), which binds to a critical DNA motif in the TCR β enhancer (Messier *et al.*, 1993). TCFβ_1 was shown to posses a DNA-binding POU-domain in its COOH-terminal region which is distantly related to other POU-domains and so may also constitute a novel class of POU-domain proteins (Messier *et al.*, 1993).

In contrast to homeodomain proteins, where the homeodomain is necessary and sufficient for DNA-binding, the POU-domain is a bipartite DNA-binding motif with both the POU_{HD} and the POU_S domain making contacts with DNA (Rosenfeld, 1991; Verrijzer *et al.*, 1992). It has been shown that the POU_{HD} and the POU_S domain are both capable, alone, of binding DNA but that both domains are required to achieve high specificity and affinity of binding (Sturm and Herr, 1988; Ingraham *et al.*, 1990; Verrijzer *et al.*, 1990, 1992). DNA binding studies and *in vitro* binding site selection studies have shown that each of the POU subdomains has a different binding specificity (Verrijzer *et al.*, 1992). The left hand of the optimal binding site is recognised by the POU_S domain and the right half by the POU_{HD} (Verrijzer *et al.*, 1992).

By the construction of chimeric proteins containing the POU_{HD} and POU_S domains of both Oct-1 and Pit-1, it has been demonstrated that both domains are involved in mediating the specificity of DNA binding (Ingraham *et al.*, 1990). Additionally, clusters of basic amino acids located at the NH_2- and COOH-terminal boundaries of the POU_{HD} and at the NH_2-terminus of the POU_S domain have also been shown to be important for high affinity binding (Sturm and Herr, 1988; Ingraham *et al.*, 1990; Treacy *et al.*, 1991).

Although Pit-1, Oct-1 and Oct-2 have been shown to behave as monomers in solution (LeBowitz *et al.*, 1989; Poellinger and Roeder, 1989; Ingraham *et al.*, 1990) and to bind to their DNA target sites as monomers, it appears that POU-domain protein-protein interactions do occur and are critical for function. Homodimerisation has been reported for Pit-1 and for Oct-2 (LeBowitz *et al.*, 1989; Poellinger *et al.*, 1989; Ingraham *et al.*, 1990), with Oct-2 co-operative interactions believed to increase transcriptional activity (Poellinger *et al.*, 1989). Heterodimer formation between different POU-domain proteins has also been reported (He *et al.*, 1989; Voss *et al.*, 1991). Such interactions are thought to be mediated through the POU_S domain and have been shown to either enhance or inhibit the action of positive POU-domain regulators. For example, the inhibitory POU-domain protein I-POU has been shown to specifically interact with the POU-domain protein cf1a, inhibiting its ability to bind and activate a specific neural promoter (Treacy *et al.*, 1991).

Interactions between POU-domain proteins and other classes of proteins have also been reported. For example, specific residues in helix 2 of the POU_{HD} of Oct-1 have been shown to be involved in mediating its ability to bind the herpes simplex virus (HSV) trans-activator VP16 (O'Hare and Goding, 1988). Additionally, the POU-domain protein Unc-86 and the homeodomain protein Mec-3 have been shown to bind to the *mec-3* promoter as a heterodimeric protein complex (Xue *et al.*, 1993). These heterodimeric interactions appear to result in positive regulation of activity but other interactions result in negative regulation. For example, protein-protein interactions between Oct-1 and the glucocorticoid receptor have been shown to result in a hormone-dependent repression of Oct-1 binding (Kutoh *et al.*, 1992). This interaction was shown to be mediated by a strong association between Oct-1 and the receptor, possibly through the POU_{HD} (Kutoh *et al.*, 1992).

Thus, in an analogous situation to many other known transcription factor families, the POU-domain may serve as both a DNA-binding domain and as a mediator of protein-protein interactions resulting in both positive and negative regulation of activity.

1.2.3.1.1.e. Zinc Finger.

Zinc fingers were first discovered in the Xenopus transcription factor TFIIIA (Brown *et al.*, 1985; Miller *et al.*, 1985) and homologous structures were subsequently discovered in a number of eukaryotic proteins, including Sp1 (Kadonaga *et al.*, 1987), and the early growth response genes, Egr-1 and Egr-2 (also referred to as NGFIA and NGFIB) (Chavier *et al.*, 1988; Christy *et al.*, 1988; Sukhatme *et al.*, 1988), where they play an important role in mediating protein-DNA interactions (Berg, 1986; Klug and Rhodes, 1987).

The TFIIIA-like zinc finger motif is usually present in tandem repeats and is characterised by the following sequence pattern, derived from the alignment of zinc finger sequences from a number of proteins:

$$Cys\text{-}X_{2\text{ or }4}\text{-}Cys\text{-}X_{12}\text{-}His\text{-}X_{3\text{-}5}\text{-}His$$

The structure of zinc fingers has been determined (Berg, 1988; Parraga *et al.*, 1988; Lee *et al.*, 1989) and studies have shown that zinc fingers contain an antiparallel β-sheet and an α-helix. Two cysteine residues near a turn in the β-sheet region and two histidines in the α-helix co-ordinate a central zinc ion and hold the structure together to form a compact globular domain.

From the crystal structure of three zinc fingers from the mouse immediate early protein zif 268 (Christy *et al.*, 1988) complexed to DNA, it has been shown that the zinc fingers bind in the major groove of DNA and wrap part way around the double helix (Pavletich and Pabo, 1991). Each finger docks against the DNA in a similar

manner, is related to the next by a simple helical motion and makes base contacts with a three base pair subsite (Nardelli *et al.*, 1991; Pavletich and Pabo, 1991).

The transcription factor Sp1 possesses three zinc fingers which are closely related to those of zif 268 (Kadonaga *et al.*, 1987) and are therefore thought to bind DNA in the same way. However, other zinc finger containing proteins, such as the human GL1 oncogene, have been shown to contact DNA in a slightly different manner with some fingers having no contacts with the DNA (Pavletich and Pabo, 1993). Therefore, it is likely, as is found for HTH motifs, that zinc fingers can dock against DNA in a number of different ways. Subtle variants are found among TFIIIA-like zinc fingers, such as the distinctive sequence patterns found in the alternating fingers of the protein ZFY (Mardon and Page, 1989), which may reflect differences in the structure or docking arrangements of the fingers (Kochoyan *et al.*, 1991).

1.2.3.1.1.f. Steroid hormone receptors.

Steroid hormone receptors belong to a large superfamily of nuclear receptors which includes the receptors for the steroid hormones, retinoids, vitamin D, thyroid hormones and several orphan receptors for which no physiological ligand is yet known (Beato, 1989). The steroid hormone receptors contain separate domains for hormone binding, DNA binding and for transcriptional activation (Beato, 1989). The DNA recognition sites of these nuclear receptors contain two copies of six base pair sequences, termed half sites, which are orientated as inverted repeats (Kumar and Chambon, 1988; Tsai *et al.*, 1989). The DNA-binding domain comprises approximately 70 amino acid residues and contains eight conserved cysteine residues which can be organised into two zinc fingers. These zinc fingers are very different in several aspects from those of TFIIIA. The zinc fingers of TFIIIA use a pair of cysteines and a pair of histidines to co-ordinate a zinc ion where as each zinc finger of steroid hormone receptors contains four cysteine residues tetrahedrally co-ordinating a zinc ion (Freedman *et al.*, 1988). Steroid hormone receptor zinc fingers also differ structurally from TFIIIA zinc fingers. NMR studies have shown that the DNA-binding domains from the glucocorticoid and estrogen receptors fold into a single globular domain consisting of a pair of α-helices positioned roughly perpendicular to each other with each binding a zinc ion (Frankel and Pabo, 1988; Hard *et al.*, 1990; Schwabe *et al.*, 1990).

From the crystal structure of the glucocorticoid receptor it has been shown that the peptide binds DNA as a dimer, with the zinc finger domains making base specific contacts with the half sites (Hard *et al.*, 1990; Luisi *et al.*, 1991). The first helix of each subunit fits into the major groove with side chains contacting the edge of base pairs. The second helix of each subunit appears to provide the dimerisation interface and also

contacts the DNA backbone. Therefore, the zinc fingers from steroid receptors in addition to playing a role in DNA binding are important mediators of dimerisation.

1.2.3.1.1.g. Leucine zipper.

The leucine zipper was first discovered as a conserved sequence motif in several eukaryotic transcription factors (Landschultz *et al.*, 1988). It is now known to be present in a wide range of proteins including the yeast transcription factor GCN_4 (Hope and Struhl, 1986; Oliphant *et al.*, 1989), the mammalian CCAAT/enhancer-binding proteins (C/EBP) (Landschulz *et al.*, 1989; Cao *et al.*, 1991; Williams *et al.*, 1991), members of the CREB/ATF (cyclic AMP element binding protein/activating transcription factor) family (reviewed in Habener, 1990), the proto-oncogene products Fos and Jun (reviewed by Morgan and Curran, 1991) and the erythroid-specific factor NF-E2 (Andrews *et al.*, 1993).

Leucine zipper (bZIP) proteins contain a DNA-binding motif of about 60 to 80 amino acids which can be subdivided into two distinct subdomains, a leucine zipper region which mediates dimerisation and a basic region which contacts the DNA (Hope and Struhl, 1986; Landschultz *et al.*, 1988).

The leucine zipper domain spans approximately 30 to 40 residues and is characterised by a heptad repeat of leucines, with a conserved repeat of hydrophobic residues often found to the NH_2-terminal side of the repeat (Landschultz *et al.*, 1988). The dimerisation regions have been shown by NMR and X-ray crystallography studies to form two parallel α-helices which form a coiled-coiled structure (O'Shea *et al.*, 1989; 1991, Saudek *et al.*, 1991).

The basic region contains about 30 amino acids and is rich in arginines, lysines and other residues conserved throughout the family (reviewed by Kerppola and Curran, 1991). Domain swap experiments have shown that this region is primarily responsible for the DNA sequence specificity of bZIP proteins (Agre *et al.*, 1989; Suckow *et al.*, 1993) and it has been proposed to form an α-helix in the dimer complex when bound to DNA (O'Neil *et al.*, 1990; Patel *et al.*, 1990; Talanian *et al.*, 1990; Weiss *et al.*, 1990).

Two closely related models have been proposed for the structure of bZIP dimers. In these models the dimer is shaped like Y in which the stem represents the leucine zipper dimerisation regions and the arms represent the basic regions of each monomer (Vinson *et al.*, 1989; O'Neil *et al.*, 1990). In the 'scissors grip' model the α-helix from the basic region is kinked so that it can follow the curve of the major groove (Vinson *et al.*, 1989) and in the 'induced helical fork' model the α-helix is straight and so extends away from the DNA after contacting three or four base pairs (O'Neil *et al.*, 1990).

The existence of the leucine zipper region in bZIP proteins potentially allows for a large degree of biological control through the formation of both homodimers and heterodimers. Dimeric complexes are known to form between Fos and Jun family members (Bohmann *et al.*, 1987), between members of the CREB/ATF family (Hai *et al.*, 1989) and between members of the C/EBP family (Roman *et al.*, 1990; Cao *et al.*, 1991; Williams *et al.*, 1991; Hsu *et al.*, 1994). An additional degree of diversity comes from the findings that proteins from the different families can also heterodimerise through their leucine zipper domains. Members from the CREB/ATF family can heterodimerise with Fos or Jun family members (Hai and Curran, 1991) and with proteins from the C/EBP family (Vallejo *et al.*, 1993). These resulting heterodimers are each thought to each posses different DNA-binding specificities and affinities and thus are expected to be targeted to different DNA sites. Heterodimer formation can also act to limit activity, as illustrated in the case of CREB whose activity is antagonised by the formation of heterodimers with CREM (Foulkes *et al.*, 1991), and C/EBPβ (or LAP) whose activity is limited by the formation of heterodimers with CHOP-10, a protein which contains a leucine zipper domain but lacks a basic region and therefore does not bind DNA (Ron and Habener, 1992). The formation of homo- and heterodimers between bZIP proteins may also be controlled by other proteins such as Tax, which transcriptionally activates the human T cell leukemia virus type-1 (HTLV-1) promoter and has been shown to act by promoting the dimerisation of bZIP proteins and so enhancing their DNA binding (Wagner and Green, 1993).

1.2.3.1.1.h. Helix-loop-helix.

Helix-loop-helix (bHLH) proteins were initially identified based on the fact that they posses a DNA-binding motif similar to that of bZIP proteins (Murre *et al.*, 1989a). Like the bZIP family, bHLH proteins contain an NH_2-terminal basic region that contacts the DNA and a neighbouring region that mediates dimer formation. However, sequence analysis has demonstrated that the dimerisation region of bHLH proteins, unlike the leucine-zipper region, forms an α-helix, a loop and a second α-helix, thus classing them as helix-loop-helix proteins (Murre et al., 1989a). The helix-loop-helix region is thought to promote dimerisation through its putative amphipathic helices in an analogous fashion to the leucine zipper region of bZIP proteins (Davis *et al.*, 1990; Voronova and Baltimore, 1990).

The bHLH family of proteins includes, among others, the muscle-determining factor MyoD and the immunoglobulin κ-chain enhancer binding proteins such as E47, E12 and E2A (Murre et al., 1989a). Like the bZIP proteins, the activity of bHLH proteins is modulated by homodimer and heterodimer formation. For example, MyoD binds DNA most tightly to its E-box DNA target site when it forms a heterodimer with

the E2A protein (Weintraub *et al.*, 1991). Other oligomerization combinations can completely disable bHLH activity. When MyoD, E12 or E47 associate with the Id group proteins, which contain a HLH region but lack a basic region, their DNA-binding and biological activities are eliminated (Benezra *et al.*, 1990; Christy *et al.*, 1991; Sun *et al.*, 1991).

A family of related proteins containing two different adjacent dimerization regions, a leucine zipper and a HLH region, termed basic-region helix-loop-helix leucine zipper (bHLH-Zip) proteins, also exists. Members of this family include the oncogene products Myc and Max (Murre *et al.*, 1989a) and the transcription factors AP4 (Hu *et al.*, 1990), USF (Gregor *et al.*, 1990), TFE3 (Beckman *et al.*, 1990) and TFEB (Carr and Sharp, 1990). Several models have been proposed for the structure of the bHLH-Zip motif (Anthony-Cahill *et al.*, 1992; Halazonetis and Kandil, 1992; Vinson and Garcia, 1992; Ferre-D'Amare *et al.*, 1993). The favoured model predicts that, in an analogous situation to bZIP proteins, the leucine zipper region forms α-helices and together with the HLH domain interacts with its partner polypeptide in the dimer with specific amino acids within the basic region mediating DNA binding (Feldmann *et al.*, 1993).

As for the bHLH proteins, the bHLH-Zip proteins can form a variety of homo- and hetero-oligomers, resulting in multiprotein complexes with diverse DNA-binding properties and biological effects. For example, Myc heterodimerises with Max to form a protein complex which will bind DNA and activate transcription (Blackwood and Eisenman, 1991; Pendergast *et al.*, 1991; Amati *et al.*, 1992; Kato *et al.*, 1992; Kretzner *et al.*, 1992), yet Myc does not dimerise with itself readily nor does it bind DNA alone (Pendergast *et al.*, 1991, Kato *et al.*, 1992, Littlewood *et al.*, 1992). Again, in an analogous situation to bHLH proteins, other oligomerization combinations can completely disable bHLH-Zip activity. For example, Max homodimers or heterodimers, formed by the interaction of Max with the Max interactor proteins Mad and Mxi (Ayer *et al.*, 1993; Zervos *et al.*, 1993), will bind Myc-Max DNA target sites with high affinity, yet Mad-Max and Max-Mxi heterodimeric complexes are not thought to activate transcription and Max-Max homodimers are thought to either repress or only weakly activate transcription (Amati *et al.*, 1992; Kato *et al.*, 1992; Kretzner *et al.*, 1992; Ayer *et al.*, 1993; Fisher et al., 1993; Zervos *et al.*, 1993).

Finally, an interesting interaction has recently been reported between a member of the bHLH family, MyoD, and the bZIP protein, Jun (Bengal *et al.*, 1992; Li *et al.*, 1992). It appears that Jun can inhibit the transactivation potential of MyoD and that this may occur through a physical interaction between the leucine-zipper region of Jun with the helix-loop-helix region of MyoD.

1.2.3.1.1.i. Helix-span-helix.

Proteins possessing helix-span-helix (HSH) motifs include the mammalian transcription factor AP-2 (Williams and Tjian, 1991). The HSH motif, sufficient for DNA binding and essential for dimerisation, has been localised to a region of 200 amino acids located towards the COOH-terminal region of the protein (Williams and Tjian, 1991). The NH_2-terminal region of the HSH motif is rich in basic residues and is predicted to bind DNA with the COOH-terminal end, potentially α-helical in nature, corresponding to the dimerisation domain. Although the overall organisation of a basic region and an adjacent dimerisation domain is similar to that of bZIP and bHLH proteins, AP-2 does not contain a leucine repeat and its dimerisation domain shares no homologies with, and is twice the length of, the corresponding domains from bZIP and bHLH proteins (Williams and Tjian, 1991).

1.2.3.1.1.j. HMG domains.

The HMG domain consists of two homologous repeats of an 80-amino-acid sequence and was first described as being present in the high mobility group proteins (HMG) (Goodwin *et al.*, 1973; Walker *et al.*, 1980).

This motif has subsequently been discovered, in one or more copies, in many DNA-binding proteins including the regulator of ribosomal RNA gene transcription, hUBF (Upstream Binding Factor) (Jantzen *et al.*, 1990) and in the transcription factor LEF-1 (lymphoid enhancer factor) (Travis *et al.*, 1991).

From the comparison of the DNA binding sites of HMG-domain proteins it appears that the structure of the DNA may be equally or more important than the actual sequence of the target site for determining the specificity of binding. HMG1 appears to bind to a DNA structure termed a four way junction which is formed by the looping of double stranded DNA to generate an X shape (Bianchi *et al.*, 1989). The structural specificity of binding was found to reside in the HMG domain and interestingly binding was found to be independent of DNA sequence (Bianchi *et al.*, 1992).

It has been suggested that HMG-domain proteins may participate in site-specific recombination reactions by bending or supercoiling DNA and so forming and sealing loops in the DNA. This has been shown to be the case for the HMG-domain protein LEF-1 which is involved in site-specific recombination reactions and bends DNA by approximately 130° (Geise *et al.*, 1992).

Such alterations in the structure of DNA have been shown to be required for achieving the correct conformation of DNA for transcription. DNA looping is a favoured model by which enhancer binding proteins interact with promoter bound proteins (Ptashne, 1986, 1988), suggesting that HMG-domain proteins may play an important role in the regulation of gene expression.

1.2.3.1.2. Activation domains.

In addition to multiple types of DNA-binding domains, proteins which regulate transcription can also possess several types of activation domains.

A majority class possesses a highly acidic activation domain (AAD), enriched in glutamate and aspartate residues, and other proteins possessdomains which are rich in glutamine or proline residues.

Additionally, some transcription factors contain domains composed of combinations of the different domains. For example, the transcriptional activation domain of c-Jun encompasses a proline-rich sequence interspersed with a negatively charged domain (Struhl, 1987) and Oct-2 possesses a domain rich in both proline and glutamine residues (Clerc *et al.*, 1988; Muller-Immergluck *et al.*, 1988; Scheidereit *et al.*, 1988; Staudt *et al.*, 1988).

1.2.3.1.2.1. Acidic transcriptional activators.

The acidic activators are the most characterised activation domains and appear unique in that they can apparently function universally in all eukaryotes tested from yeast to man.

Acidic activators, including the yeast proteins GCN_4 and GAL4 and the HSV virion protein VP16, have commonly been assumed to posses two features critical for function, an acidic region and a structure which is either α-helical or unstructured in nature.

The importance of acidic regions for mediating transcriptional activation in the yeast proteins GCN_4 and GAL4 was demonstrated in a series of early experiments. Such experiments included deletion analysis of the activation region of GCN_4 which demonstrated that the transcriptional activity of the protein is generally correlated with the number of negative charges (Hope *et al.*, 1988). Other experiments demonstrated that the activation of transcription by chimeric proteins formed by the fusion of the GAL4 DNA-binding domain to random *E. coli* DNA sequences required no apparent sequence specificity except for a net negative charge (Ma and Ptashne, 1987a).

Acidic transcriptional activators have been proposed to adopt either an amphipathic α-helical structure or to exist as unstructured 'acid blobs' (Giniger and Ptashne, 1987; Ma and Ptashne, 1987a; Ptashne, 1988; Sigler *et al.*, 1988).

The possible involvement of an α-helix for mediating the interaction between AADs and their target factor(s) was demonstrated in a series of genetic studies. When the DNA-binding domain of GCN_4 was fused to a DNA sequence predicted to fold into an amphipathic α-helix with a negatively charged face, a high level of activated transcription was detected (Giniger and Ptashne, 1987). The importance of the acidic

residues residing on one face of the putative α-helix was suggested from the finding that the GAL4 DNA-binding domain, when fused to a sequence in which the acidic residues would no longer be on one face, was no longer able to activate transcription (Giniger and Ptashne, 1987).

Sigler (1988) proposed an alternative structure for AADs in which they exist as unstructured 'acid blobs' which interact with their target factor(s) through purely ionic interactions. The findings that progressive deletions of the GCN_4 AAD resulted in a step-wise rather than a dramatic loss of transcriptional activation (Hope *et al.*, 1988) and that random segments of DNA encoded activator functions (Ma and Ptashne, 1987a) prompted this hypothesis. The notion of an unstructured activation domain was further supported by studies showing that VP16 is largely devoid of a secondary structure in solution (Donaldson and Capone, 1992; O'Hare and Williams, 1992).

Recently, these accepted models concerning the structure and function of acidic transcriptional activators have been questioned. Firstly, there is no direct evidence that the activation domains are indeed α-helical and, if so, if this structure is important for transcription and secondly, point mutations in non-acidic residues of VP16 and GAL4 can result in significantly reduced activation potentials (Cress and Triezenberg, 1991; Leuther *et al.*, 1993). Additionally, several studies have demonstrated that there is no strict correlation between negative charge and activation potential (Ma and Ptashne, 1987b; Hope *et al.*, 1988; Leuther *et al.*, 1993). An alternative model predicts that the acidic amino acids in such activators are not, as commonly thought, required for function and that this region is not unstructured or α-helical, but that it may adopt a β-sheet structure (Leuther *et al.*, 1993; Van Hoy *et al.*, 1993).

Leuther *et al.* (1993) studied the COOH-terminus of the GAL4 protein which is bifunctional, containing both an AAD and a region that binds the negative regulator GAL80. Mutational analysis allowed the selection of mutants in which one function was maintained while the other one was altered (Leuther *et al.*, 1993). From such analysis it was revealed that all of the acidic residues in the COOH-terminus are dispensable for activation and that mutants with no charge, or indeed an overall positive charge, are capable of activating transcription at near wild-type levels. It was concluded, from this finding, that it is not the charge of amino acids that is important for activation but rather the hydrophobicity of the amino acids which interact with the target factor(s). Results from such genetic studies also led to the suggestion that the COOH-terminal region of GAL4 forms an antiparallel β-sheet structure as deletion of a region predicted to form part of the β-sheet was found to almost completely abolish activation by GAL4 (Leuther *et al.*, 1993). It was concluded from these studies that the

face of the β-sheet containing most of the hydrophobic residues interacts with the target factor(s) while the other side may form the binding surface for the GAL80 protein.

In a separate study, Van Hoy *et al.* (1993) carried out a biophysical analysis of peptides corresponding to the activation domains of GAL4 and GCN_4. The results from this study also demonstrated that the peptides are not α-helical in structure, but that in slightly acidic solutions they adopt an extensive β-sheet structure.

It is not yet clear what role, if any, the β-sheet structure of such proteins plays in mediating transcriptional activation. More detailed genetical and physical studies will be necessary in order to further understand the structure and function of such activation domains.

1.2.3.1.2.2. Glutamine-rich transcriptional activators.

Sp1 is a widely characterised sequence-specific DNA-binding protein which possesses an activation domain rich in glutamine residues (Courey and Tjian, 1988; Kadonaga *et al.*, 1988). Other transcriptional activators containing glutamine-rich domains include the product of the *Drosophila zeste* gene (Biggin *et al.*, 1988) and the products of a number of homeo-box containing genes such as *Antp* (Schneuwly *et al.*, 1986).

Courey and Tjian (1988) used cells derived from *Drosophila* embryos in order to carry out a functional analysis of human Sp1. From the analysis of a variety of deletion mutants of Sp1 four distinct regions outside the DNA-binding domain were identified which are capable of modulating transcriptional activity. Two of the identified domains, shown to be functionally redundant of one another, correspond to glutamine-rich segments in the NH_2-terminal portion of the protein. Unlike acidic activation domains, no relationship was found between charge and activity but some kind of higher order structure, in addition to glutamine richness, was thought to be required for high levels of Sp1 activity (Courey and Tjian, 1988). In addition to the glutamine rich domains a highly charged segment, also distinct from acidic activation domains, was shown to be important for Sp1 transcriptional activity, although it was a great deal less active than the glutamine rich domains and a domain located at the extreme COOH-terminal of Sp1 was also found to be important for influencing transcriptional activity, in the context of the intact protein (Courey and Tjian, 1988).

1.2.3.1.2.3. Proline-rich transcriptional activators.

Transcription factors which posses activation domains rich in proline residues include members of the C/EBP family (Mermod *et al.*, 1989), proteins involved in the developmental control of gene expression in *Drosophila*, such as the *fushi tarazu* (*ftz*) and *Kruppel* gene products (Laughon and Scott, 1984; Rosenberg *et al.*, 1986), steroid hormone receptors (Gronemeyer *et al.*, 1987; Kumar *et al.*, 1987) and the transcription factor AP-2 (Williams *et al.*, 1988).

Proteins belonging to the C/EBP family have the unusual property of regulating both transcriptional initiation and DNA replication (Jones *et al.*, 1987). The biochemical properties of mutant C/EBP polypeptides have been investigated and this has allowed the identification of a transcriptional activation domain located in the COOH-terminal region of such proteins which is rich in proline residues (Mermod *et al.*, 1989). Although all known members of the C/EBP family contain similar proline-rich domains, the overall sequence varies between different members (Gill *et al.*, 1988; Paonessa *et al.*, 1988; Santoro *et al.*, 1988). This proline-rich segment has been shown to be distinct and separate from the DNA-binding and replication domains and has also been shown to be distinct from either acidic or glutamine-rich activation domains.

1.2.3.2 Mechanisms of action of transcriptional activators.

A major question which has yet to be unequivocally resolved is how transcriptional activators communicate with an often distant RNA polymerase II to provoke transcriptional initiation. A favoured model proposes that, by looping out the intervening DNA, activators contact a factor in the transcriptional basic machinery at the minimal promoter (Ptashne, 1986, 1988). In support of such a model, evidence has been presented suggesting that transcriptional activators can bind to either the TATA-binding protein TBP, its associated complex TFIID (Section 1.2.2.2.a.) and/or the general transcription factor TFIIB (Section 1.2.2.2.c.).

Analyses with affinity chromatography have pointed to TBP as a potential target for VP16 and for Tat, a potent transactivator of HIV transcripton (Stringer *et al.*, 1990; Kashanchi *et al.*, 1994). Additionally, mutational analysis has shown that VP16 molecules which have mutations introduced into their activation regions are defective in binding to TBP, suggesting that VP16 TBP protein-protein interactions are important for mediating transcriptional activation (Ingles *et al.*, 1991).

Other lines of evidence point to both TBP and TAFs, associated with TBP in the TFIID complex, having a role to play as potential targets. This idea came from the findings that TBP itself can mediate only basal levels of transcription (Hoffmann *et al.*, 1990; Peterson *et al.*, 1990) and that the TAFs are required for mediating activated levels (Sawadogo and Roeder, 1985b; Horikoshi *et al.*, 1988; Nakajima *et al.*, 1988;

Dynlacht *et al.*, 1991; Tanese *et al.*, 1991). For example, it has been shown that the AAD of VP16 requires the presence of other components, in addition to TBP, present in crude TFIID fractions to activate *in vitro* transcription in HeLa cell extracts (White *et al.*, 1991, 1992). It has also been demonstrated that stimulation of *in vitro* transcription in HeLa cells by a variety of chimeric activators, containing distinct activation domains, requires the presence of both TBP and unique TAFs (Brou *et al.* 1993). Additionally, it has been shown that the activation domain of the adenovirus large E1A protein specifically binds to Holo-TFIID (Boyer and Berk, 1993). It is thought that TAFs associate with the conserved COOH-terminal domain of TBP to mediate transcriptional activation as Holo-TFIID assembled with an NH_2-terminal deleted form of TBP fully supports transcriptional activation (Zhou *et al.*, 1993).

TFIIB is also thought to be a target for transcriptional activators (Lin and Green, 1991). As for TBP, it has been shown that the AAD of VP16 can selectively interact with TFIIB and it also been shown that, under some conditions, VP16 must be present during incorporation of TFIIB into the preinitiation complex to allow for activated levels of transcription (Lin and Green, 1991). Additionally, VP16 activation mutants, in an analogous situation to TBP, are deficient in binding to TFIIB (Lin *et al.*, 1991). Further evidence to support the suggestion that protein-protein interactions between TFIIB and activators are important for mediating transcriptional activation comes from the findings of Roberts *et al.* (1993). Mutagenesis of the putative amphipathic helix in TFIIB, which is necessary for interaction with VP16, was shown to abolish the ability of TFIIB to both bind VP16 and to respond to transcriptional activation. However, TFIIB was still able to function in basal transcription, providing strong evidence that TFIIB plays a role in both basal and activated transcription. Evidence also exists for interactions occurring between TFIIB and other classes of activator proteins. Recently, an interaction between *Drosophila* TFIIB and the glutamine-rich activation domain of the *fushi tarazu* (*ftz*) gene product was demonstrated in *Drosophila* cells (Colgan *et al.*, 1993) and a separate study demonstrated a specific interaction between human thyroid hormone receptor β (hTRβ) and TFIIB (Baniahmad *et al.*, 1993). hTRβ, when localised in the nucleus, is a DNA-binding protein capable of both transcriptional silencing and hormone-dependent activation of transcription, with the ligand binding domain (LBD) thought to be required for the silencing function (Baniahmad *et al.*, 1992). It has been shown that the LBD binds TFIIB and that thyroid hormone, which converts the receptor from a silencer to an enhancer, decreases the interaction of the LBD with TFIIB. Thus, it has been suggested that TFIIB, through binding hTRβ, may mediate its ability to silence transcription (Baniahmad *et al.*, 1993). Therefore, it has been shown that TFIIB may be capable of mediating both activation and silencing of transcription.

A popular model to explain how activators involve TBP, TFIIB and TAFs in the activation of transcription initiation predicts that they facilitate their binding to the promoter (Lin and Green, 1991; Choy and Green, 1993). Lin and Green (1991) demonstrated that acidic activators enhance the stable entry of TFIIB into the preinitiation complex, thus increasing a rate limiting step in preinitiation complex assembly. These studies were further extended to shown that TFIIB interacts with both TFIID and activator proteins and that this interaction increases the binding of TFIIB to the preinitiation complex (Choy and Green, 1993). This step alone was shown to be insufficient for transcriptional activation and a second step, occurring after TFIIB entry, was shown to be affected by transcriptional activators. It is proposed that activators may also contact a general transcription factor(s) which assembles after TFIIB or that they may work through TAFs to affect the assembly of other general transcription factors (Choy and Green, 1993).

Another model predicts that activators interact with the general transcription factors to somehow convert them from an inactive to an active form (Ouzounis and Sander, 1992). In this model the factors are thought to assemble non-productively at the core promoter, in the absence of activators, to form inactive preinitiation complexes. In the presence of activators a productive initiation complex may form so allowing for high levels of transcription. Baniahmad *et al.* (1993) used this model to explain the observed transcriptional silencing mediated through TFIIB. It is postulated that, in the absence of thyroid hormone, hTRβ stabilises an inactive form of TFIIB in the preinitiation complex so precluding it from association with other basal transcription factors. In the presence of hormone however, the interaction of hTRβ with TFIIB is blocked and therefore TFIIB is relieved from repression and is subsequently accessible for the activation of transcription (Baniahmad *et al.*, 1993).

An alternative, but not mutually exclusive, model to explain how activators facilitate transcription initiation suggests that their binding to the general transcription factors acts to compete the binding of factors that repress transcription (reviewed by Hahn, 1993). Several factors have been shown to bind TBP and to repress transcription, including Dr1, Dr2 (Inostroza *et al.*, 1992) and Negative Cofactors 1 and 2 (NC1 and NC2) (Meisterernst and Roeder, 1991; Meisterernst *et al.*, 1991). Negative factors that bind TFIIB may also exist and it is thought that activators may directly compete with such negative factors for binding to TBP and/or TFIIB and so increase the rate of transcriptional initiation (Hahn, 1993).

Chromatin-associated proteins, such as nucleosomes and histone H1, are also thought to play an important role in response to activators. Nucleosomes consist of 160-200 bp of DNA wrapped around an octamer core of histones, bound at the beginning, the centre and the end by a single molecule of histone H1. The nucleosome

is a highly stable structure and presents a formidable obstacle to transcription (reviewed by Svaren and Chalkley, 1990). Depletion of nucleosomes *in vivo* has been shown to dramatically increase basal transcription (Durrin *et al.*, 1992) and it has been suggested that transcriptional activators disrupt chromatin structure in a mechanism independent of replication (reviewed by Svaren and Chalkley, 1990). Proposed models predict that cofactors or mediators somehow act to facilitate the interaction of transcriptional activators with the promoter by the modification of chromatin structure (reviewed by Adams and Workman, 1993). For example, the glucocorticoid receptor is thought to mediate the effects of hormonal induction on the mouse mammary tumour virus (MMTV) promoter by displacement of histone H1 and disruption of the nucleosome core, thereby allowing nuclear factor 1 (NF1) and Oct-1 to bind to their target DNA motifs (Bresnick *et al.*, 1992; Lee and Archer, 1994).

Therefore, there are several models to explain how transcriptional activators act to provoke transcriptional initiation at RNA polymerase II promoters. The elucidation of the exact mechanisms by which different activators achieve their end results have yet to be fully resolved but it is likely that several will exist including those mentioned above.

In conclusion, it is clear that many different classes of transcription factors exist possessing the ability to both specifically recognise specific DNA sequence elements and to regulate the transcriptional machinery. The aim of this study was to identify and characterise transcription factors binding sites within a region of the rPPT-A promoter which potentially may act to regulate the transcriptional activity of the rPPT-A gene.

Section 2: Materials and Methods.

2.1. Materials.

2.1.1. Commonly used solutions.

5 x Agarose gel loading buffer:	0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 1mM EDTA, 30% glycerol.
MOPS:	41.8g/l MOPS, 4.1g/l NaOAc, 3.4 g/l EDTA.
10 x PBS:	1.37M NaCl, 27mM KCl, 80mM Na_2HPO_4, 15mM KH_2PO_4.
10 x TBE:	pH8.3; 108.9 g/l Tris base, 55.7 g/l boric acid, 4.7 g/l EDTA. pH8.8; 162 g/l Tris Base, 27.5 g/l boric acid, 9.5 g/l EDTA.
TE:	10mM Tris HCl (pH8), 1mM EDTA.

Tris-saturated phenol: Frozen phenol was thawed by bringing to room temperature and then heating in a water bath to 68°C. 0.1 % hydroxyquinoline was added as a preservative and phenol was extracted with an equal volume of 0.5M Tris HCl (pH 8.0) and then extracted with equal volumes of 0.1M Tris HCl (pH 8.0) until buffer pH was >7.0. Phenol was finally extracted with 0.1 vol of 0.1M Tris HCl (pH8.0) and stored at 4°C under a layer of 0.1M Tris HCl (pH8.0).

Phenol-chloroform: Tris-saturated phenol as prepared above was mixed with an equal volume of chloroform.

2.1.2. Bacterial culture medium.

L-Broth:	10g/l tryptone, 5g/l yeast extract, 5g/l NaCl.
L-Agar:	L-broth to which 1% bacto-agar was added.

Ampicillin: A stock of ampicillin was made up at 100mg/ml and stored at 20°C.

L-Agar plates: L-agar was melted in a microwave oven and left to cool to room temperature. Ampicillin was added to the agar at a final concentration of 0.1 mg/ml before pouring into plates and leaving to set at room temperature. Once set, plates were used or stored at 4°C for up to one week.

2.1.3. Plasmids.

A series of plasmids were supplied by Dr. Karen Chapman (Department of Medicine, Western General Hospital, Edinburgh). These contain the following fragments from the rPPT-A promoter linked to a *lac* Z (β-galactosidase) reporter gene (Mulderry *et al.*, 1993), where +1 corresponds to the first nucleotide of the major transcriptional initiation site (identified by Carter and Krause, 1990):

prPPT-βGal3: -856 to +92
prPPT-βGal4: -671 to +92
prPPT-βGal6: -345 to +92
pVL29: -431 to +92

These plasmids were all sequenced (Section 2.2.7.1.) to determine accurately which regions of the rPPT-A promoter they contain.

The plasmid pPPT was supplied by Dr. J. Quinn (MRC Brain Metabolism Unit, Edinburgh). It contains a fragment from the rPPT-A promoter, extending from -3356 to +443, cloned into the vector pUC19 (Quinn, 1992).

2.2. Methods.

2.2.1. Gel electrophoresis.

2.2.1.1. Horizontal agarose gel electrophoresis.

The percentage of agarose used depended on the size range of linear DNA to be separated. 1% (w/v) agarose (Agarose MP, Boehinger) was used to separate DNA fragments of ~7 kb to 0.56 kb, 1.5% to 2% (w/v) agarose was used to separate DNA

fragments of ~1 kb to 0.1 kb and 5% (w/v) agarose (NuSeive GTG) was used to separate DNA fragments or oligonucleotides of < 0.1 kb.

100ml horizontal gel slabs (140mm x 110mm x 30mm) containing the correct percentage of agarose were placed in horizontal gel electrophoresis apparatus (Gibco BRL, Model HS) containing sufficient 1 x TBE buffer to cover the gel. Samples were loaded in 0.2 vol loading buffer and gels were usually electrophoresed at 60mA for as long as required to see sufficient separation of DNA fragments. After electrophoresis, gels were stained in 1 x TBE buffer containing ethidium bromide (0.5µg/ml) for 5 mins and DNA fragments were visualised using a long wave UV transilluminator. DNA was photographed using Asa Polaroid film type 667 in a Polaroid Cu-5 camera.

25ml mini agarose gels (83mm x 57mm x 30mm) (Gibco BRL, Model Horizon 58) were also used in some cases to analyse DNA. Typically, these gels were electrophoresed in 1 x TBE at 60mA until sufficient separation of DNA fragments had occurred.

2.2.1.2. Vertical gel electrophoresis.

a. 4%/6% Native polyacrylamide gel electrophoresis.

4%/6% native polyacrylamide gels were prepared by mixing 4ml/6ml of 50% 29:1 bis:acrylamide (Biorad, premixed) with 2.5ml of 10 x TBE in a final volume of 50ml. Gels were cross-linked with 50µl of TEMED, catalysed by 0.5ml of 10% APS and immediately poured into 150mm x 170mm x 1.5mm glass sandwiches for use in electrophoretic mobility shift assays (Section 2.2.6.1.), 300mm x 400mm x 0.4mm glass sandwiches for use in exonuclease protection assays (Section 2.2.6.3.) or 170mm x 880mm x 0.4mm glass sandwiches for use in DNase 1 footprinting assays (Section 2.2.6.2.), previously sealed with 1% (w/v) agarose. 20-tooth combs (1.5mm) were placed in the top of the gels, between the plates, and the gels were then left to polymerise. Polymerised gels for electrophoretic mobility shift assays were attached to vertical gel electrophoresis apparatus (Gibco BRL, Model V16-2) containing 0.5 x TBE buffer and pre-electrophoresed for 30 mins at 100V. Polymerised gels for use in exonuclease protection assays were attached to vertical gel electrophoresis apparatus (Gibco BRL, Model S2) containing 1 x TBE buffer and pre-electrophoresed for 1 h at 1000V. Polymerised gels for use in DNase 1 footprinting assays were attached to vertical gel electrophoresis apparatus (Gibco BRL, Model SA) containing 1 x TBE buffer and pre-electrophoresed for 1 h at 1000V. Samples were loaded while the gels were still running. Electrophoretic mobility shift assays were electrophoresed for a further 1-2 h at 200V, exonuclease protection assays were electrophoresed for 3-4 h at 1500V and DNase 1 footprinting assays were electrophoresed for a further 16 h at

1500V. Following electrophoresis, gels were transferred to Whatman No.1 paper and dried using a heated vacuum gel drier (Biorad Model S83). Radiolabelled samples were visualised by exposing the dried gels to autoradiographic film for 1-24 h (Section 2.2.1.3.).

b. 6% Denaturing polyacrylamide/urea gel electrophoresis.

6% denaturing polyacrylamide gels were prepared by dissolving 42g of urea (Aristar, BDH), 20ml of 30% (29:1) bis:acrylamide (Biorad, premixed) and 10ml of 10 x TBE in a final volume of 100ml. The polyacrylamide solutions were filtered though Whatman membrane filters (0.45µm) under vacuum and degassed before adding 0.55ml of 10% APS and 55µl of TEMED. Solutions were immediately poured into 300mm x 400mm x 0.4mm glass sandwiches previously sealed with gel sealing tape (Gibco BRL). 20-tooth combs (0.4mm) were inserted and the gels were left to polymerise. After polymerisation, gels were attached to vertical gel electrophoresis apparatus (Gibco BRL, Model S2) containing 1 x TBE buffer and pre-electrophoresed at 1000V for 30 mins prior to loading samples. Electrophoresis was then carried out for approximately 2-3 h. Following electrophoresis, gels were transferred to Whatman No.1 paper covered with Saran wrap and dried using a heated vacuum gel drier (Biorad Model 583). Radiolabelled samples were visualised by autoradiography (Section 2.2.1.3.).

2.2.1.3. Autoradiography.

Detection of radiolabelled material was carried out at -70°C using Fuji RX X-ray film (18 x 24cm or 30 x 40cm). X-ray film was placed against dried gels inside autoradiographic cassettes with 2 x or 4 x intensifying screens (in the case of material containing ^{32}P). After the appropriate exposure time films were developed for 2 mins in Kodak LX-24 developer (0.25 dilution), rinsed in water, fixed for 1 min in Kodak FX-40 fixer (0.25 dilution), rinsed once more in water and dried.

2.2.2. Cloning and DNA techniques.

2.2.2.1. Plasmid construction.

pSM1/Q5 and pSM1'/Q5 were constructed as follows: 10µg of plasmid pPPT was digested with *Xba* I and *Eco*R I (Section 2.2.2.2.). Reaction products were electrophoresed through a 1% low melting point agarose gel (Section 2.2.1.1.) and the DNA fragment of 3.9 Kb in size, containing rPPT-A DNA sequence, was purified (Section 2.2.2.3.a.). Purified DNA was then digested with *Hinf* I which generated 8 fragments, the desired fragment of 604 bp in length was again purified from a low

melting point agarose gel. Sticky ends were filled in with 2mM dCTP, dGTP, dATP, dTTP, 1 unit of Klenow (large fragment of DNA polymerase I) and 2µl 10 x Klenow buffer (0.5M Tris HCl pH7.6, 0.1M $MgCl_2$, 10mM DTT) in a final volume of 20µl and incubated at room temperature for 15 mins.

2µg of pUC19 vector was digested with *Sma* I and then dephosphorylated by adding 1 unit of calf intestinal phosphatase (CIP) and incubating at 37°C for 1 h. Dephosphorylated vector and filled in insert were then ligated (Section 2.2.2.5.) and competent cells were transformed with the ligation mix (Section 2.2.2.6.). The transformed ligation mix was spread out on L-agar plates containing ampicillin plus X-gal and IPTG to allow for blue/white selection. Restriction analysis of DNA minipreps (Section 2.2.2.7.) confirmed insertion of the rPPT-A DNA fragment in both orientations. Large scale plasmid preparations (Section 2.2.2.8.) were then carried out and the insert sequence was confirmed by DNA sequencing (Section 2.2.7.1.).

2.2.2.2. Restriction enzyme digests.

DNA was digested typically using 1 unit of restriction enzyme per 1µg of DNA, 1 x reaction buffer (as specified by the manufacturer) and dH_2O to the required volume. Reaction mixtures were generally incubated at 37°C for 1 h. Digested DNA was visualised after electrophoresis though an agarose gel (Section 2.2.1.1.). Diagnostic restriction enzyme digests generally used 0.2 - 0.5µg of DNA in a final volume of 10µl. Preparative restriction enzyme digests used increasing amounts of DNA in a final volume of 50µl. Restricted DNAs were then either purified using the Magic™ DNA clean up system (Promega) or desired fragments were recovered after electrophoresis of the entire 50µl digests though low melting point agarose gels (Section 2.2.2.3.).

DNA size markers routinely used were *Hind* III fragments of λ which give a range of DNA fragment sizes between 0.56 and 23 kb or *Hinf* I fragments of pAT153 which give a range of sizes between 75 bp and 1.631 kb.

2.2.2.3. Purifying DNA from low melting point agarose gels.

a. DNA fragments of > 0.1 kb.

DNA fragments were electrophoresed through low melting point agarose gels, as described (Section 2.2.2.1.). After visualising on a UV transilluminator, gel slices containing the DNA fragments of interest were excised using a clean scalpel blade. Gel slices were placed in 1.5ml Eppendorf tubes and heated to 65-70°C for 10-15 mins or until the agarose was completely melted. An equal volume of Tris-saturated phenol was added and the tubes were vortexed for 10 secs before incubating in a dry ice bath for 5-

15 mins. This was followed by a 10 min spin in an Eppendorf bench top centrifuge after which top aqueous layers were transferred to new eppendorf tubes. The aqueous layers were then extracted once more with an equal volume of Tris-saturated phenol, once with equal volumes of Tris-saturated phenol and chloroform:isoamylalcohol (24:1, v/v) and finally with an equal volume of chloroform:isoamylalcohol (24:1, v/v). Each extraction involved a 10 sec vortex followed by a 2 min centrifugation in an Eppendorf bench top centrifuge. DNA fragments were then precipitated with 2.5 vol of ethanol and 0.15% NaCl (Section 2.2.2.4.).

b. DNA fragments of < 0.1 kb and oligonucleotides.

DNA fragments or oligonucleotides were electrophoresed through 5% NuSeive agarose gels (Section 2.2.1.1.) and recovered using the phenol extraction protocol, according to manufacturers instructions.

After visualising on a UV transilluminator, gel slices containing the DNA fragments were excised using a clean scalpel blade. Gel slices were placed in 1.5 ml eppendorf tubes and melted at 67°C for 10 mins. 4 to 5 vol of TE (warmed to 67°C) were added to melted gel slices and mixed well. An equal volume of Tris-saturated phenol was then added to diluted agarose slices, mixed for 15 secs and centrifuged for 3 mins in an Eppendorf bench top centrifuge. Aqueous top layers were then extracted once more with an equal volume of Tris-saturated phenol followed by extraction with an equal volume of phenol/chloroform and finally with an equal volume of chloroform:isoamylalcohol (24:1, v/v). DNA present in the aqueous layers was then ethanol precipitated with 0.1 vol of 3M NaOAc and 2 vol of 95% ethanol (cooled to -20°C) (Section 2.2.2.4.). After 30 mins at -20°C, reactions were centrifuged in an eppendorf bench top centrifuge at room temperature for 30 mins. Pellets were washed with 2 vol of 80% ethanol (cooled to -20°C), stored in a dry ice bath for 15 mins and re-centrifuged as above. Pellets were dried under vacuum in a Gyrovap for 10 mins or until they were dry and were finally resuspended in an appropriate volume of TE or dH_2O.

2.2.2.4. Ethanol precipitation of DNA.

2-2.5 vol of ethanol and either 0.5 vol of 7.5M NH_4Ac or 0.1 vol of 3M NaOAc were added to DNA samples and they were then vortexed for 10 secs. This was followed by a 5-10 min incubation in a dry ice bath or incubation at -20°C for 30 mins to overnight and then a 10 min centrifugation in an Eppendorf bench top centrifuge. DNA pellets were dried at room temperature for up to 30 mins and then resuspended in an appropriate volume of TE or H_2O.

2.2.2.5 DNA ligation.

Ligation reactions were carried out in a final volume of 10µl containing 1µl of 10 x ligation buffer (0.5M Tris HCl pH7.6, 100mM $MgCl_2$, 10mM DTT), 1mM ATP, 1 unit of T4 DNA ligase (BCL) and a 1:5 molar ratio of vector DNA:insert DNA (typically 10-50ng of vector DNA and 50-250ng of insert DNA). Ligation reactions were incubated at room temperature for 3-4 h.

2.2.2.6. Preparation and transformation of competent *E. coli* .

A single colony of *Escherichia coli* strain HB101 was grown overnight in 3ml of L-broth, diluted into 40ml of L-broth and grown in an orbital shaker at 37°C to an A_{600} of 0.4-0.6. Cells were harvested by centrifugation at 6000rpm in a Sorvall SS34 rotor for 5 mins at 4°C. Cell pellets were placed on ice, resuspended in 0.4 vol of ice cold 0.1M $MgCl_2$ and pelleted as before. The pellets were then resuspended in 0.05 vol of ice cold 0.1M $CaCl_2$ and left on ice for 45 mins. The cells were pelleted as before, resuspended in 1 vol of a buffer containing 0.1M MOPS pH6.5, 50mM $CaCl_2$, 20% glycerol and left on ice for a further 20 mins to acquire full competence. Competent cells were aliquoted into 1ml eppendorf tubes, snap frozen in a dry ice bath and stored at -70°C.

1-5µl of ligation mix or 10ng of plasmid DNA was added to 100-200µl of competent HB101 cells and incubated on ice for 30 mins. The mixture was then heat shocked at 45°C for 2 mins and returned to ice for 2 mins. The cells were added to 1ml of L-broth and incubated at 37°C for 1 h. They were then briefly pelleted, resuspended in 100µl of L-broth and subsequently spread on L-agar plates containing 0.1mg/ml ampicillin which had previously been spread with 40µl of 2% X-gal (diluted in ethanol) and 40µl of 100mM IPTG. The plates were incubated overnight at 37°C and colonies which grew on the selective media and appeared white were picked for mini-prep analysis (2.2.2.7.).

2.2.2.7. Mini-prep analysis of transformed *E.coli* colonies.

Mini-prep DNA was prepared by the alkaline lysis method. Colonies were picked from agar plates into 3ml of L-broth containing 0.05mg/ml ampicillin and grown overnight at 37°C in a rotator. The following day cells from 1ml of the overnight culture were placed in an eppendorf tube and pelleted in a bench top centrifuge for 1 min (the remaining 2ml of culture was stored at 4°C for use in large scale preparation of plasmid DNA Section 2.2.2.8.). Pellets were resuspended in 100µl of ice cold GTE buffer (50mM glucose, 25mM Tris HCl pH8.0, 10mM EDTA) and mixed by vortexing. 200µl of freshly prepared alkaline-SDS solution (0.2M NaOH, 1% (v/v) SDS) was then added, gently mixed and tubes were placed on ice. 150µl of ice cold

potassium acetate (3.0M KAc, 11.5% (v/v) glacial acetic acid) was added followed by a 5 min incubation on ice. Samples were then centrifuged for 5 mins and supernatants were added to fresh eppendorf tubes and extracted with 0.5 vol of Tris-saturated phenol and 0.5 vol of chloroform:isoamyl alcohol (24:1, v/v). The aqueous phases were precipitated (Section 2.2.2.4.) with 2 vol of ethanol at room temperature for 5 mins. After centrifugation for 5 mins in an Eppendorf bench top centrifuge, the pellets were dried at 37°C and resuspended in 50µl of TE containing 20µg/ml of DNase-free pancreatic RNase (Sigma) (prepared by boiling a stock of 1mg/ml for 10 mins, aliquoting into eppendorfs and storing at -20°C). Mini-prep DNA was stored at -20°C and 5-10µl was used for restriction analysis.

2.2.2.8. Large scale preparation of plasmid DNA.

2ml of overnight cultures, containing individual colonies picked from agar plates, were diluted into 300-500ml of L-broth, containing 0.05mg/ml ampicillin and grown overnight in an orbital shaker at 37°C. The following day cultures were pelleted by centrifugation at 6000rpm for 5 mins at 4°C in a Sorvall RC-5B centrifuge using a GSA rotor. After decanting the supernatant, bacterial pellets were resuspended in 12ml of ice-cold GTE buffer (50mM glucose, 25mM Tris HCl pH8.0, 10mM EDTA) containing approximately 1mg/ml lysozyme. 24ml of freshly prepared alkaline-SDS (0.2M NaOH, 1% (v/v) SDS) was then added and suspensions were left on ice for 5 mins. 16ml of ice cold potassium acetate (3.0M KAc, 11.5% (v/v) glacial acetic acid) was added and suspensions were incubated for a further 10 mins on ice. This was followed by centrifugation at 4000rpm for 15 mins in a Sorvall GSA rotor at 4°C to pellet out cell debris and cellular DNA. Supernatants containing the plasmid DNA were strained though butter muslin into 250ml centrifuge pots and precipitated with 0.6 vol of isopropanol at room temperature for 30 mins. Plasmid DNA was recovered by centrifugation at 5000rpm for 15 mins at room temperature in a Sorvall GSA rotor. Supernatants were drained away and the pellets were washed with 70% ethanol, dried and resuspended in 1ml of TE. The DNA samples were then brought to a total volume of 4ml with TE after which 400µl of ethidium bromide (10mg/ml) was added. Exactly 4.4g of CsCl was then dissolved into the solutions followed by a 5-10 min incubation on ice. The solutions were centrifuged in 15ml polpropylene tubes (Corning) at 5000rpm for 5 mins at room temperature. Supernatants were transferred to 6ml polyallamer ultracentrifuge tubes (Du Pont) and topped up with a stock solution containing 1.1g of CsCl in 1.0ml of H_2O and 0.1 ml of EtBr (10mg/ml). Ultracentrifuge tubes were then sealed with an ultracrimp tool and centrifuged in a Sorval TV-1665 ultracentrifuge rotor at 45000rpm for 16 h at 20°C. Supercoiled plasmid DNA bands were visualised by daylight or a long wave UV transilluminator

and recovered with a large bore needle and syringe. Ethidium bromide was removed from the plasmid DNA by extracting several times with equal volumes of butan-1-ol. The plasmid preparations were then dialysed for 24 h against three 2 litre changes of TE at 4°C. The concentration of recovered plasmids was determined by spectrophotometry (OD_{260} 1.0 = 50μg DNA/ml).

2.2.3. Tissue culture growth conditions.

Tissue culture work was performed by E. Clark and C. Morrison. HeLa cells (Human cervix carcinoma), AtT20 cells (mouse pituitary tumour) and C1300 cells (mouse neuroblastoma) were grown on 75cm^2 tissue culture flasks (J. Bibby Science Products) in DMEM (Dulbeccos Modified Eagles Medium; Flow Laboratories) supplemented with 10% fetal calf serum (Globepharm), 2mM glutamine (Gibco BRL), 100 units/ml penicillin and 100μg/ml streptomycin (Gibco BRL).

PC12 cells (rat adrenal pheochromocytoma) were grown on collagen coated 75cm^2 tissue culture flasks (J.Bibby Science Products) in DMEM, 5% fetal calf serum, 10% horse serum, 100 units/ml penicillin and 100μg/ml streptomycin.

2.2.4. Protein techniques.

2.2.4.1. Preparation of whole cell extracts.

C1300, AtT20 and PC12 cells were cultured as described (Section 2.2.3.). Whole cell extracts were prepared essentially as described by Manley *et al.* (1980). Approximately 2 x 10^6 cells were harvested by scraping the surface of the culture flasks. The cell pellets were washed once in ice cold 1 x PBS and transferred to eppendorf tubes. All subsequent steps were carried out at 4°C.

Cells were resuspended in 1.5ml of 1 x buffer A (10mM HEPES, 10mM KCl, 1mM $MgCl_2$, with protease inhibitors 0.25mM DTT and 0.5mM PMSF, 0.25M stock made up in isopropanol, added just before use) and were left to swell for 8 mins on ice. Cells were then pelleted by centrifugation in a bench top eppendorf centrifuge for 2 mins at 4°C. Cell pellets were resuspended in an equal volume of 1 x buffer C (20mM HEPES, 20% glycerol, 0.42M NaCl, 1.5mM $MgCl_2$, 0.2mM EDTA with 0.25mM DTT and 0.5mM PMSF added just before use) and lysed with 2 strokes of a 1.5ml plastic homogeniser. The suspensions were left on ice for 20 mins vortexing every 5 mins. The cell lysates were finally centrifuged in a bench top eppendorf centrifuge for 2 mins at 4°C and the supernatants were carefully decanted into fresh eppendorf tubes, snap frozen in a dry ice bath and stored at -70°C.

2.2.4.2. Preparation of nuclear extracts.

HeLa cells were cultured as described (Section 2.2.3.). Nuclear extracts were prepared essentially as described by Dignam *et al.* (1983). Approximately 6 x 10^6 cells were harvested by centrifugation at 2500rpm for 10 mins at 4°C in a bench top centrifuge (Chilspin, MSE). All subsequent steps were carried out at 4°C.

The cell pellets were washed in 3 vol of ice cold 1 x PBS and collected again by centrifugation at 2000rpm for 10 mins at 4°C in a bench top centrifuge (Chilpsin, MSE). Cells were then swollen by resuspending in 3 vols of ice cold 1 x buffer A and were left to stand on ice for 10 mins. Swollen cells were collected by centrifugation at 2000rpm for 10 mins at 4°C in a bench top centrifuge (Chilpsin, MSE). Cells were resuspended in 2 vol of ice cold 1 x buffer A and lysed under ice with 10 strokes of a dounce homogeniser (B pestle). The cells were then pelleted by centrifugation at 3000rpm for 10 mins at 4°C in a Sorvall GSA rotor and the supernatants containing unwanted cellular membranes and nucleic acid material were removed. The pellets were centrifuged once more at 15000rpm for 25 mins at 4°C in a Sorvall GSA rotor and the supernatants removed. Pellets were resuspended in 0.05 vol of 1 x buffer C, lysed with 10 strokes of a dounce homogeniser (B pestle) and left to stand on ice for 20 mins. The nuclear samples were then centrifuged at 38000rpm for 45 mins at 4°C in a Sorval TV-1665 ultracentrifuge rotor and the supernatants collected. This material was dialysed against three 1 litre changes of 1 x 100mM buffer D (20mM HEPES, 20% glycerol, 0.1M KCl, 0.2mM EDTA) and finally clarified by centrifugation in a bench top eppendorf centrifuge for 10 mins at 4°C. The supernatants containing nuclear material were then aliquoted into fresh eppendorf tubes and snap frozen.

Nuclear extracts were prepared from rat brain tissue essentially as above after being dissected from rat adult male Cob Wistars (by Mr. R.C. Dow, MRC, Brain Metabolism Unit, 1 George Square, University of Edinburgh).

2.2.4.3. Determination of protein concentration.

Protein concentrations of whole cell and nuclear extracts were determined using the Biorad protein assay. Protein standard solution (BSA) was diluted to a 20µl final volume in dH_2O to give concentrations of 1.0, 0.5, 0.25, 0.125, 0.0625 and 0.031 mg/ml. Each dilution, plus a 20µl dH_2O control, was mixed well with 1ml of Biorad protein assay solution, previously diluted 1 part in 5 with dH_2O. Protein standards were then transferred to plastic cuvettes and the A_{595} of each was obtained using a Shimadzu UV/60 spectrophotometer to produce a standard calibration curve.

An appropriate volume of protein extract (typically 1-5µl) was then mixed with 1ml of the diluted Biorad protein assay solution and the A_{595} was measured as above.

Protein extract concentrations were calculated by relating the A_{595} of protein extract to the standard curve.

2.2.5. Radioactive labelling of DNA.

2.2.5.1. Preparation of oligonucleotide probes.

Oligonucleotides (Oswell DNA services, Edinburgh), for use in electrophoretic mobility shift assays, were labelled either at the 5' end with T4 polynucleotide kinase or at the 3' end using the Klenow fragment of *E. coli* DNA polymerase. In some cases oligonucleotides were purified from 5% NuSeive agarose gels before use (Section 2.2.2.3.b.).

a. 5' End labelling.

Double stranded oligonucleotides (100ng) were 5' end labelled at room temperature using 15µCi of [γ^{32}P] ATP (Amersham, 3000Ci/mmol), 10 units of T4 polynucleotide kinase and 1 x kinase buffer (10 x; 1M Tris HCl pH8.0, 0.1M $MgCl_2$, 100mM DTT). After 30 mins dH_2O was added to a total volume of 50µl and the reactions were terminated by extraction with an equal volume of Tris-saturated phenol. Labelled oligonucleotides were separated from unincorporated label by centrifugation though Nu-Clean sephadex D25 disposable spun columns (IBI) and the total volume was brought up to 100µl with dH_2O.

b. 3' End labelling.

Double stranded oligonucleotides (100ng) (with TCGA or TA overhangs) were filled in at room temperature using 25µCi of [α^{32}P] dTTP (Amersham, 3000Ci/mmol), 5 units of the Klenow fragment of *E.coli* DNA polymerase and 1 x Klenow buffer (10 x; 0.5M Tris HCl pH7.6, 0.1M $MgCl_2$, 10mM DTT). Octamer oligonucleotide (with a GAG overhang) was filled in as above using 25µCi of [α^{32}P] dCTP in the place of dTTP. After 20 mins dH_2O was added to a total volume of 50µl and the reactions were terminated by extraction with an equal volume of Tris-saturated phenol. Labelled oligonucleotides were separated from unincorporated label by centrifugation though Nu-Clean sephadex D25 disposable spun columns (IBI) and the total volume was brought up to 100µl with dH_2O.

2.2.5.2. Labelling DNA probes for DNase 1 footprinting assays.

The plasmids used in DNase 1 assays are listed in materials (Section 2.1.3.). The method used to generate suitable substrates for DNase 1 footprinting assays are outlined in Figure 3.

Typically, 10µg of each substrate plasmid was linearised by digestion in the polylinker with 10 units of *Sal* I (Section 2.2.2.2.). Enzyme activity was terminated by passing linearised plasmids over Magic™ DNA clean up system columns (Promega). The 5' ends were dephosphorylated with CIP (24 units, at 37°C for 1 h). The CIP was inactivated using Magic™ DNA clean up system columns and linearised dephosphorylated plasmids were subsequently labelled at the 5' ends with T4 polynucleotide kinase. 1µg of plasmid was incubated with 50µCi of [γ^{32}P] ATP (Amersham, 3000Ci/mmol), 10 units of T4 polynucleotide kinase and 1 x kinase buffer at room temperature for 30 mins. Labelled DNA was seperated from unincorporated label using the Magic™ DNA clean up system.

Labelled linearised plasmids were then cut with a second restriction enzyme in order to generate two singly end labelled fragments, one of almost full length and the other of 10-20 bp in length which would not interfere with the DNase 1 assay. The enzymes chosen produced DNA fragments in which the region of promoter to be analysed was located approximately 50-200 bp from the labelled end. Finally, enzyme activity was terminated using the Magic™ DNA clean up system and the DNA fragments were eluted in 30µl of TE.

Another approach involved restricting the plasmid pSM1'/Q5 with *Eco*R I and *Hind* III, separating the digestion products on a 1% low melting point agarose gel and purifying the desired DNA fragment of approximately 650 bp, containing rPPT-A promoter sequence and pUC19 polylinker sequence (Section 2.2.2.3.a.). The purified DNA was then restricted with *Dde* I, again the digestion products were separated on a 1% low melting point agarose gel and the desired DNA fragment of approximately 500 bp was purified. This purified DNA fragment was dephosphorylated and 5' end labelled as described above. Finally, it was restricted with *Sal* I, passed over a Magic™ DNA clean up system column and was ready for use in DNase 1 footprinting assays.

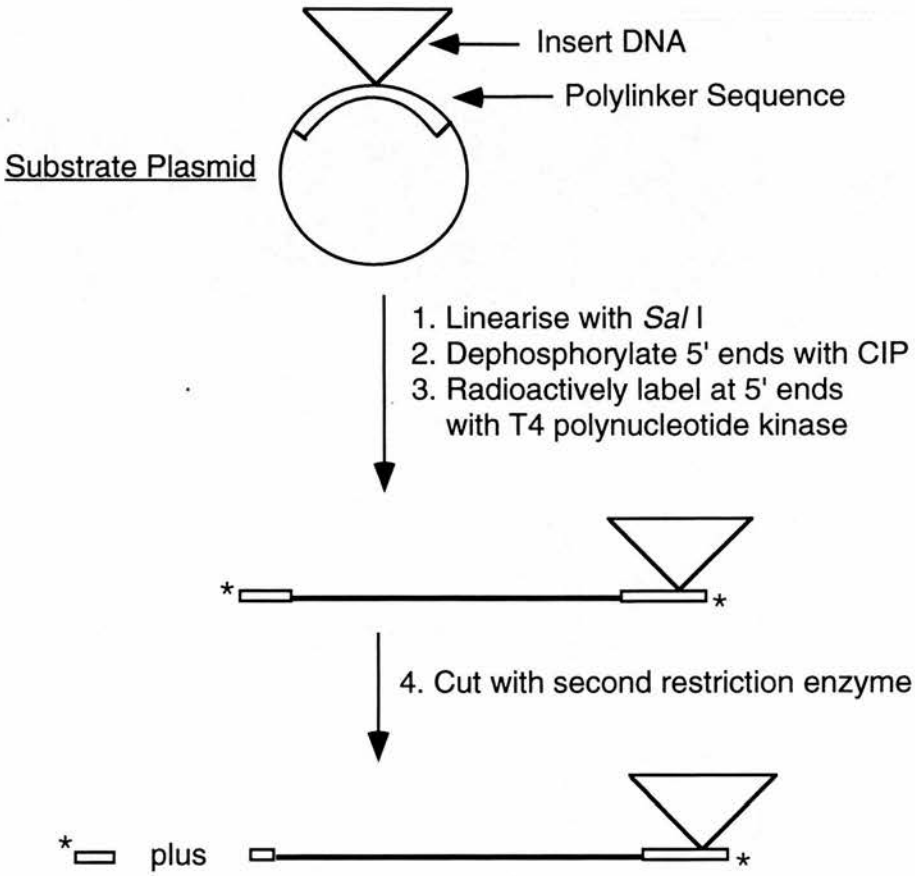
2.2.5.3. Labelling DNA probes for exonuclease assays.

10µg of the plasmid pSM1/Q5 was restricted with *Pvu* II (Section 2.2.2.2.), digestion products were separated on a 1% low melting point agarose gel and the DNA fragment containing rPPT-A promoter sequence plus the *lac* operator sequence from

Figure 3. Diagrammatic representation of the method used to generate suitable substrates for DNase 1 footprinting assays.

The plasmid shown (substrate plasmid) represents the constructs prPPT-βGal3, prPPT-βGal4, prPPT-βGal6, pVL29, pSM1/Q5 and pSM1'/Q5, with the sites of polylinker sequence and insert DNA sequence noted.

1. The plasmids were digested in the polylinker with the restriction enzyme *Sal* I.
2. The 5' ends of the resulting linearised plasmids were then dephosphorylated using CIP.
3. Dephosphorylated, linearised plasmids were then radioactively 5' end labelled with T4 polynucleotide kinase.
4. This was followed by digestion with a second restriction enzyme to remove 10-20 bp of DNA sequence from one end of the linearised, labelled DNA fragments. This resulted in a singly end labelled substrate DNA fragment for use in DNase 1 footprinting analysis.


Insert DNA
Polylinker Sequence
Substrate Plasmid
1. Linearise with *Sal* I
2. Dephosphorylate 5' ends with CIP
3. Radioactively label at 5' ends
with T4 polynucleotide kinase
4. Cut with second restriction enzyme
plus

pUC19 was purified (Section 2.2.2.3.a.). The resulting blunt ended fragment was then 3' end labelled using the Klenow fragment of *E. coli* DNA polymerase as follows.

1µg of the DNA fragment was initially incubated with 5 units of the Klenow fragment of *E.coli* DNA polymerase and 1 x Klenow buffer at room temperature for 15 mins. This produced 5' overhanging ends due to the 3' to 5' exonuclease activity of the Klenow fragment. The resulting 5' overhanging ends were then filled in using the 5' to 3' polymerase activity of the Klenow fragment as follows. The DNA fragment was incubated with 40µCi of [α^{32}P] dATP, dCTP, dGTP and dTTP (Amersham, 800Ci/mmol), 5 units of the Klenow fragment of *E.coli* DNA polymerase and 1 x Klenow buffer at room temperature for 30 mins. This labelling step was followed by a 10 min chase using 1mM of all four cold dNTPs. The Klenow fragment was then inactivated by extraction with Tris-saturated phenol and the labelled fragment was seperated from unincorporated label by centrifugation though Nu-Clean sephadex D50 disposable spun columns (IBI). The eluted labelled fragment was finally ethanol precipitated (Section 2.2.2.4.) and was ready for use in exonuclease assays.

2.2.5.4. Labelling DNA markers.

100ng of 123 bp ladder DNA (BRL) was 5' end labelled essentially following the method for 5' end labelling of oligonucleotides (Section 2.2.5.1.a.), except that Nu-Clean D50 sephadex disposable spun columns (IBI) were used to remove unincorporated label.

2.2.6. Analysis of protein/DNA interactions.

2.2.6.1. Electrophoretic mobility shift assays.

Electrophoretic mobility shift assays, as outlined below, were based on the method of Singh *et al.* (1986) with a few modifications.

A typical electrophoretic mobility shift assay contained approximately 1ng of double-stranded end labelled oligonucleotide (Section 2.2.5.1.), 0.5-2.5µg (depending on batch lot) of non-specific competitor poly [d(I-C)] (Boehinger), 10-30µg of protein extract and sufficient 0mM or 100mM buffer D to bring the final reaction concentration to 100mM salt, in a volume of 5-10µl.

Protein extracts (nuclear or whole cell, Section 2.2.4.1. and 2.2.4.2.) were mixed with buffer D and poly [d(I-C)] and left at room temperature for up to 5 mins. Labelled oligonucleotide probe and 2µl of dye mix (2% (w/v) bromophenol blue, 2% (w/v) xylene cyanol) were then added and the reaction mixtures were incubated at room temperature for 20 mins. Where appropriate, excess unlabelled competitor

oligonucleotide was included in the reaction mixture. Competitors were added along with labelled oligonucleotide probe, unless stated, at 10 to 250 molar excess over ^{32}P labelled oligonucleotide. Reactions were then loaded onto 4% or 6% native polyacrylamide gels (Section 2.2.1.2.a.) and protein/DNA complexes were resolved from unbound DNA by electrophoresing at 200V for 1-2 h. Dried gels were exposed to autoradiographic film for 1-16 h (Section 2.2.1.3.).

2.2.6.2. DNase 1 footprinting assays.

Singly end labelled DNA fragments (Section 2.2.5.2.) (100ng per reaction) were mixed with assay buffer (40mM Tris HCl pH7.5, 6mM $MgCl_2$) and brought to a total volume of 20µl with dH_2O. DNA fragments were then mixed with approximately 1mg of protein extract (whole cell or nuclear, Section 2.2.4.1. and 2.2.4.2.) which had been preincubated with 5 to 30µg of poly [d(I-C)] for 2 mins at room temperature. Reaction mixtures were incubated at room temperature for 20 mins and were then digested with DNase 1 enzyme (1-4µl) (Pharmacia FPLC pure, 7.5 units/µl) at 37°C for 2 mins. The amount of DNase 1 enzyme added was predetermined in preliminary experiments. DNase 1 enzyme activity was terminated by the addition of an equal volume of 1 x stop buffer (10mM Tris HCl pH7.4, 25mM EDTA, 0.5% SDS, 50mM NaOAc) followed by extraction with an equal volume of phenol/chloroform. Finally, reactions were ethanol precipitated (Section 2.2.2.4.) and resuspended in 4µl of stop solution (95% deionised formamide, 20mM EDTA, 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanol). Reactions were boiled for 2 mins then loaded onto 6% denaturing polyacrylamide gels (170mm x 880mm x 0.4mm) (Section 2.2.1.2.a.), along with ^{32}P labelled marker DNA (Section 2.2.5.4.) and a Maxam-Gilbert sequencing reaction of the labelled fragment (Section 2.2.7.2.). Gels were electrophoresed for approximately 16 h at 1500V and once dried, were exposed to autoradiographic film (Section 2.2.1.3.), typically for 16 h.

2.2.6.3. Exonuclease protection assays.

The exonuclease protection assay is a method which allows for the identification of sequence specific protein binding sites within a particular fragment of DNA. This method makes use of immunobeads, consisting of anti-rabbit Ig coupled to anti-β-galactosidase, in order to concentrate specifically bound proteins. These beads are prepared as outlined below.

a. Preparation of 10 x immunobeads.

Immunobeads carrying goat anti-rabbit Ig coupled to anti-β-galactosidase antibodies were prepared following the method of Levens and Howley (1985).

Immunobeads (Biorad) were mixed with anti-rabbit-β-galactosidase antibody at a volume ratio of 11:1 in the presence of 10μg/ml BSA as a carrier to prevent non-specific binding. This reaction mixture was incubated at room temperature for 3 h inverting every 20 mins and the beads were recovered by centrifugation at 3000rpm for 10 mins at room temperature in a bench top centrifuge (Chilpsin, MSE). Pelleted beads were washed twice in 1 x binding buffer (10x; 0.1M Tris HCl pH7.5, 0.25M NaCl, 10mM DTT, 75mM $MgCl_2$), re-pelleted and finally resuspended in 10 x binding buffer plus 1mM DTT and 100μg/ml BSA to produce 10 x binding beads.

b. Exonuclease protection assay.

Exonuclease protection assays were carried as previously described (Quinn *et al.*, 1987).

^{32}P-3'-end-labelled DNA fragments (Section 2.2.5.3.) (100ng per reaction) were immobilized on immunobeads via a *lac* repressor-β-galactosidase fusion protein as follows. DNA was mixed with 0.1μg of *lac* repressor-β-galactosidase fusion protein (Promega) and 1 x binding buffer in a total volume of 10μl. Reactions were left at room temperature for 10 mins to allow the *lac* repressor protein to bind to the *lac* operon of pUC19. A predetermined volume (typically 15-25μl) of 10 x immunobeads (Section 2.2.6.3.a.) was then added and reactions were left at room temperature for a further 15 mins to allow the beads to bind the β-galactosidase portion of the protein. Beads were recovered by centrifugation in a bench top eppendorf centrifuge for 30 secs then washed in 1 x 100mM D buffer, re-pelleted and resuspended in 10μl (per reaction) of 1 x exonuclease buffer (10mM Tris HCl pH8.0, 5% glycerol, 4mM $MgCl_2$, 0.2mM DTT).

10μl (per reaction) of ^{32}P-3'-end-labelled DNA linked to immunobeads was incubated with approximately 0.2mg of protein extract (whole cell or nuclear, Sections 2.2.4.1. and 2.2.4.2.) in a final volume of 15μl, including 100mM 1 x buffer D and 2-5μg of poly [d(I-C)], at room temperature for 30 mins. Immunobead-DNA-protein complexes were recovered by centrifugation in a bench top eppendorf centrifuge for 30 secs and resuspended in 15μl of 1 x exonuclease buffer. 40 to 80 units of T7 gene 6 exonuclease enzyme (USB) was then added in order to digest the DNA from the 5' end to the 3' end. Reactions were incubated at room temperature for 3 mins and terminated by the addition of an equal volume of 1 x stop buffer. Reaction volumes were brought

to 30µl with dH_2O and then extracted with an equal volume Tris-saturated phenol. Finally, reactions were ethanol precipitated (Section 2.2.2.4.), resuspended in 4µl of stop solution and electrophoresed through 6% denaturing polyacrylamide gels (300mm x 400mm x 0.4mm) along with ^{32}P labelled marker DNA (Section 2.2.5.4.) at 1500V for 3-4 h. Dried gels were exposed to autoradiographic film (Section 2.2.1.3.), typically for 16 h.

2.2.7. DNA sequencing.

2.2.7.1. Dideoxy sequencing.

Plasmid DNA was sequenced using Sequenase version 2.0 according to manufacturers instructions (USB).

DNA samples (2-5 µg) were denatured in a final volume of 20µl containing 4µl of 1M NaOH and incubated at room temperature for 5 mins. Denatured samples were then ethanol precipitated in 8µl of 5M NH_4Ac and 100µl of EtOH (Section 2.2.2.4.) and resuspended in 5µl of dH_2O. The DNA samples were then annealed to the appropriate primer in a final volume of 10µl containing 2µl of 5x sequenase buffer (5x; 200mM Tris HCl pH 7.5, 100mM $MgCl_2$, 250mM NaCl), 5µl of denatured DNA and 5ng of primer. For sequencing the plasmids pSM1/Q5 and pSM1'/Q5 the primer supplied with the Sequenase version 2.0 kit was used (termed -40; 5' GTTTTCCCAGTCACGAC 3'). For sequencing the plasmids prPPT-βGal3, prPPT-βGal4, prPPT-βGal6 and pVL29 a primer containing sequence from the rPPT-A promoter, spanning nucleotides +4 to -20 relative to the major transcriptional start site, was used (sequence; 5' TCGACTCTAGTCCCTGCTCCTGCTTCGC 3'). Reactions were boiled for 2 mins and then slowly cooled to room temperature. Annealed DNA samples (10µl) were then radioactively labelled with ^{35}S dATP by incubation with 1µl of 0.1M DTT, 2µl of labelling mix (7.5µM dCTP, 7.5µM dGTP, 7.5µM dTTP, diluted 1:5 in dH_2O), 5µCi of [^{35}S] dATPαS (Amersham, 1000Ci/mmol) and 2µl of Sequenase Version 2.0 (diluted 1:8 in enzyme dilution buffer: 10mM Tris HCl pH7.5, 5mM DTT, 0.5mg/ml BSA) at room temperature for 2-5 mins. Reactions were terminated by the incubation of 3.5µl of each labelled DNA sample with 2.5µl aliquots of all four ddNTPs (prewarmed to 37°C) followed by an incubation at room temperature for 5 mins. 4µl of stop solution was then added and samples were boiled for 2 mins before loading onto 6% denaturing polyacrylamide/urea gels (Section 2.2.1.2.b.). Gels were electrophoresed at 1500V for 3-4 h and fixed for 30 mins (5% methanol, 15% acetic acid) before drying and exposing to autoradiographic film (Section 2.2.1.3.), typically for 16 h.

2.2.7.2. Maxam Gilbert sequencing.

Singly ^{32}P-5'-end-labelled DNA fragments (Section 2.2.5.2.) were sequenced using the G+A reaction in the Sigma Maxam-Gilbert kit as according to manufacturers instructions (Sigma).

Briefly, 5-10µl of DNA was depurinated by incubation with 2µl of pyridinium formate at 37°C for 30 mins. Samples were dried under vacuum using a Gyrovap, resuspended in 20µl of dH_2O and dried once more. DNA was cleaved at the modified bases by dissolving in 100µl of piperidine (diluted 1:10 in dH_2O), followed by an incubation at 90°C for 30 mins. Samples were dried once more under vacuum and resuspended in 10µl of dH_2O. Resuspended DNA samples were dried once again, resuspended in 10µl of dH_2O and finally dried and resuspended in 5µl of stop solution. Sequenced DNAs were boiled for 2 mins and quickly cooled in ice water before electrophoresing on 6% denaturing polyacrylamide gels (Section 2.2.1.2.a.), along with DNase 1 footprinting reactions (Section 2.2.6.2.).

3.1. Aim of the study.

The aim of this study was to investigate how the rat preprotachykinin-A (rPPT-A) gene is regulated at the transcriptional level.

The regulation of the initiation of mRNA synthesis involves the specific interaction of several different types of trans-acting factors with a number of cis-acting DNA sequence elements, including upstream activating sequences and enhancers, located both upstream and downstream from the major transcriptional start site (Section 1.2.3.). In order to characterise regions of DNA sequence within the rPPT-A promoter which have the potential to act as such regulatory elements, an attempt was made to identify protein/DNA interaction sites within a 1300 bp region of the promoter.

The sequence of the region of the promoter examined, spanning nucleotides -865 to +447, is shown in Figure 4 with the first nucleotide (+1) of the major transcription initiation site (identified by Carter and Krause, 1990) as an adenine base.

3.2. Sequence analysis of the rPPT-A promoter.

Two previous studies have identified putative transcription factor binding sites within the 5' flanking region of the rPPT-A promoter. Carter and Krause (1990) sequenced 866 bp of 5' flanking sequence and, by comparison with the consensus sequence for various transcription factors, identified several putative regulatory elements. These include a region homologous to an estrogen response element (ERE) at -149 to -137 (..GGTTACCGTCTCG..) (Maurer and Notides, 1987), a serum response element (SRE) at -224 to -206 (..GATGGCGAGACCTCGACTT..) (Triesman, 1986), cAMP response elements (CRE) at -41 to -30 (..ATCACGCCTGAG..) and -409 to -398 (..TTTGGCTGTCCT..) (Roesler *et al.*, 1988), a CAAT box at -181 to -177 (..CCAAT..) (Santoro *et al.*, 1988), and GC boxes at -108 to -100 (..GGGGGCGGC..) and -270 to -261 (..GGGGCGTCAG..) (reviewed by Dynan and Tjian, 1985; Briggs *et al.*, 1986). Chapman *et al.* (1993) extended this analysis and determined the nucleotide sequence of 3350 bp of 5' flanking sequence. By inspection of the sequence several putative regulatory elements were noted within the 5' flanking DNA. These sequence elements include putative octamer binding protein binding sites at -2822 to -2815 (..ATTTGTAT..), -2782 to -2775 (..ATACAAAT..) and -662 to -655 (..ATGTAAAA..) (reviewed by Schaffner, 1989) and putative AP-1 binding sites at -2743 to -2737 (..TGAGTCA..), -2382 to -2376 (..TGACTAA..), -341 to -335 (..TGAGTCA..) and -318 to --312 (..TGAGTAA..) (reviewed by Curran and Franza, 1988).

Figure 4. Nucleotide sequence of the rPPT-A promoter spanning nucleotides -865 to +447.

The nucleotide sequence is numbered on the left hand side with the first nucleotide (+1) of the major transcriptional start site as an adenine base (shown in bold type) (identified by Carter and Krause, 1990).

-865 CTGCAGAGCT CCAAAGGTAA GCATCCAGCC TTTCTAGTCC CCCAACAAGG

-815 CTAAAGGGGA GAGAGGCACA ATTATCCTCT TCCCACCCCT TCTGCCTTCA

-765 GGGTGTGCCT GGGAAGAAGC TGTAGGGGAA CAAAAGATGC CTTAGAATGG

-715 CTGATGGGTA AGTTCTACAT GAGAAAGGAG GTTTAAATTC CTCTTTCCCC

-665 TAAATGTAAA ACAAACCTGC CTTCATCCTC TGAAGCGGGA GACCGGAAAC

-615 ACTTTTGCAG TGCTAGAGAA ATGAGAATAT TCTGACTGAT TTGGTGGGGA

-565 GGGGGGTTGG GGGGGTGTGT TCCAGCCCTA GATATAACAC CTCATAAACC

-515 TTAAGACACA TAAAGTAGAA ATGAAAGGAA AACCCCGCTT GCTTCATCCC

-465 TCTGAAGTGC TTGCTGGTGT CTTAGTATTA TTCACAAGGT TTTGCTGCTC

-415 AAGTTATTTG GCTGTCCTCA AAGCGCAATA TTCCCTGATG CCTCTTGAGA

-365 GAAAAGTTCC CTAAGTCCGA AGCATGAGTC ACTTCGCTCA GTTTTGATGA

-315 GTAATCTCAG GTGTCACTGA ACCTTGTTCG GAAGAAGAGG GGAGGGGGGC

-265 GTCAGATTTG CAGACGGAAG AAAACAGGTC TCTCTGGATT GGATGGCGAG

-215 ACCTCGACTT CCCTAAAATT GCGTCATTTC GAACCCAATT TGGTCCAGAT

-165 GTTATGGACT CCGACGGGTT ACCGTCTCGG AAACTCTATC ACGCAAGCAA

-115 AAGGCGAGGG GGCGGCTAAT TAAATATTGA GCAGAAAGTC GCGTGGGGAG

-65 AGTGTCACGT GGCTCTCCAG GCTCATCACG CCTGAGATAA ATAAGGCGAA

-15 GCAGGAGCAG GGACT**A**GAGC GCACTCGGAC CAGCTCCACT CCAGCACCGC

+36 GGCGGAGGAG AGCGAGGAGC GCCCAGCAAG TGCGCACCTG CGGAGCATCA

+86 CCGGGTCCGA CCGCAGTGAG TACCACTCCC GCCTGCCATC TGCCGCTGCC

+136 TGGGGCCCGT CGTGCCTTGG TCAGAGGTGG CGTCCGCCCG GGGTTTCACC

+186 TGCAGCAGCC ACAGACCCGG AGAGAAACTG TGCAAATACC AACACCTCTC

+236 TTTGTCTGCC TTCAGGCTTC GGAGTGTGGG TCAGTGGGTA GGGGGCTGGG

+286 ACGTTGAGAG GCAAAGAGAG GAGGACTTGA GGCTCTTTGG CACGTCAGTA

+336 GCCTTCTTAA AGGGTTTGGA GAAATGTTCG ACCTGGTGGG TTTTTGTTTT

+386 TGTTTTAATC TTGGTTTAGC GAGACCTCTT CTTTCTTCCT TGTGGCATTG

+436 AAATTCAGCT GC

3.3. Functional analysis of the rPPT-A promoter.

Understanding the transcriptional regulation of the rPPT-A promoter has been hampered by the lack of cell lines which express the gene at high levels, either endogenously or when transfected with plasmids containing regions of the promoter linked to reporter genes (Wood *et al.*, 1990; Gilchrist *et al.*, 1991; and personal communications, J.Quinn).

Recently, Mulderry *et al.* (1993) demonstrated, by microinjection, that plasmids containing 3.3 Kb of rPPT-A 5'-flanking sequences linked to a *lac*Z reporter gene are expressed in cultured DRG neurons. By 5' deletion analysis, they showed that expression was directed by 865 bp of the promoter lying immediately upstream of the transcriptional start site. Furthermore, inclusion of sequences downstream of the major transcriptional start site to +526, incorporating the first intron, increased expression two to three fold.

3.4. Strategy.

The findings above suggest that many of the elements regulating rPPT-A gene expression in DRG neurons lie in 865 bp of sequence immediately 5' to the major transcriptional start site and in 3' sequence encompassing the first intron.

Based on these findings it was decided, in this study, to analyse a corresponding region of the promoter, spanning nucleotides -865 to +447 bp relative to the major transcriptional start site, for potential cis-acting DNA sequence elements. In order to identify sequences of DNA within the promoter important for function, a series of assays were used dedicated to identifying protein/DNA interaction sites. These are;

1. Exonuclease protection analysis.
2. DNase 1 footprinting analysis.
3. Electrophoretic mobility shift assays.

Unfortunately, it is very difficult to obtain material of suitable quantity and quality from DRG neurons making this tissue unsuitable for use in the types of biochemical assays that this study requires. As there are no cell lines which express the rPPT-A gene at high levels (Wood *et al.*, 1990), the well characterised HeLa cell line was used as a source of protein extract. HeLa cells have been shown to express a large array of transcription factors and many of these, including for example AP-2 (Mitchell *et al.*, 1991) and Oct-1 (Wood *et al.*, 1992), are also expressed in DRG neurons, thus making them suitable for use in this study.

Initially, a region of the rPPT-A promoter close to the transcriptional start site and TATA box was analysed for the presence of protein/DNA interaction sites using the exonuclease protection assay. This was followed by DNase 1 footprinting and

electrophoretic mobility shift analysis of a larger region of the promoter.

3.5. Analysis of the rPPT-A promoter by the exonuclease protection assay.

3.5.1. The exonuclease protection assay.

The exonuclease protection assay (Quinn *et al.*, 1987) (Section 2.2.6.3.) is a method which allows for the enrichment, identification and characterisation of sequence-specific DNA-binding proteins (Figure 5). The advantage of this assay over the more commonly used DNase 1 footprinting assay is that it does not require binding sites on the DNA to be saturated. Thus, this assay requires less protein extract than DNase 1 footprinting assays and may detect DNA-binding proteins which are less abundant and would not be identified by footprinting analysis .

The principles behind the exonuclease protection assay are outlined in Figure 5. The DNA sequence of interest is cloned into polylinker sequence of the pUC19 vector adjacent to the *lac* operator sequence (O_1 and O_3). A *Pvu* II fragment including the DNA sequence of interest and the *lac* operator sequence is radioactively 3' end labelled. This end labelled fragment is then incubated with a *lac* repressor-β-galactosidase fusion protein allowing the *lac* repressor portion of the fusion protein to bind to the operator sequence. Immunobeads, consisting of anti-rabbit Ig coupled to anti-β-galactosidase antibodies, are then added to the reaction. The β-galactosidase portion of the fusion protein will bind to the anti-β-galactosidase antibody. This results in a complex of DNA linked to immunobeads via the *lac* repressor-β-galactosidase fusion protein. The complex is then incubated with protein extract in the presence of the non-specific competitor DNA, poly(d[I-C]), to allow sequence specific proteins to bind to any recognition sites present within the DNA. Pelleting the beads by brief centrifugation concentrates the specifically bound protein(s), leaving un-bound proteins in the supernatant. T7 gene 6 exonuclease enzyme, which digests the DNA in a 5' to 3' direction, is then added to the washed pellet. In the absence of specific binding factors, digestion proceeds along the DNA only stopping at the *lac* operator-*lac* repressor interaction. On electrophoresis, this is revealed as two bands, one large (indicated as a) and one small (indicated as b). However, if factor(s) are specifically bound to a particular region of the DNA, exonuclease digestion stops at the position(s) of binding and new band(s) (indicated as b + c) appear on electrophoresis corresponding to bound factor(s) (indicated as Protein Binding Site). This is accompanied by a diminution of the intensity of the lower *lac* operator-*lac* repressor band. From the position of the protein binding site on an autoradiogram relative to size marker DNA, electrophoresed along with the exonuclease enzyme digestion products, the location of the protein binding site on the DNA can be estimated.

Figure 5. Schematic representation of the exonuclease assay.

A DNA fragment containing the sequence of interest, (), cloned next to the *lac* operator sequence (O_3 and O_1) from pUC 19 is radioactively labelled at the 3' ends.

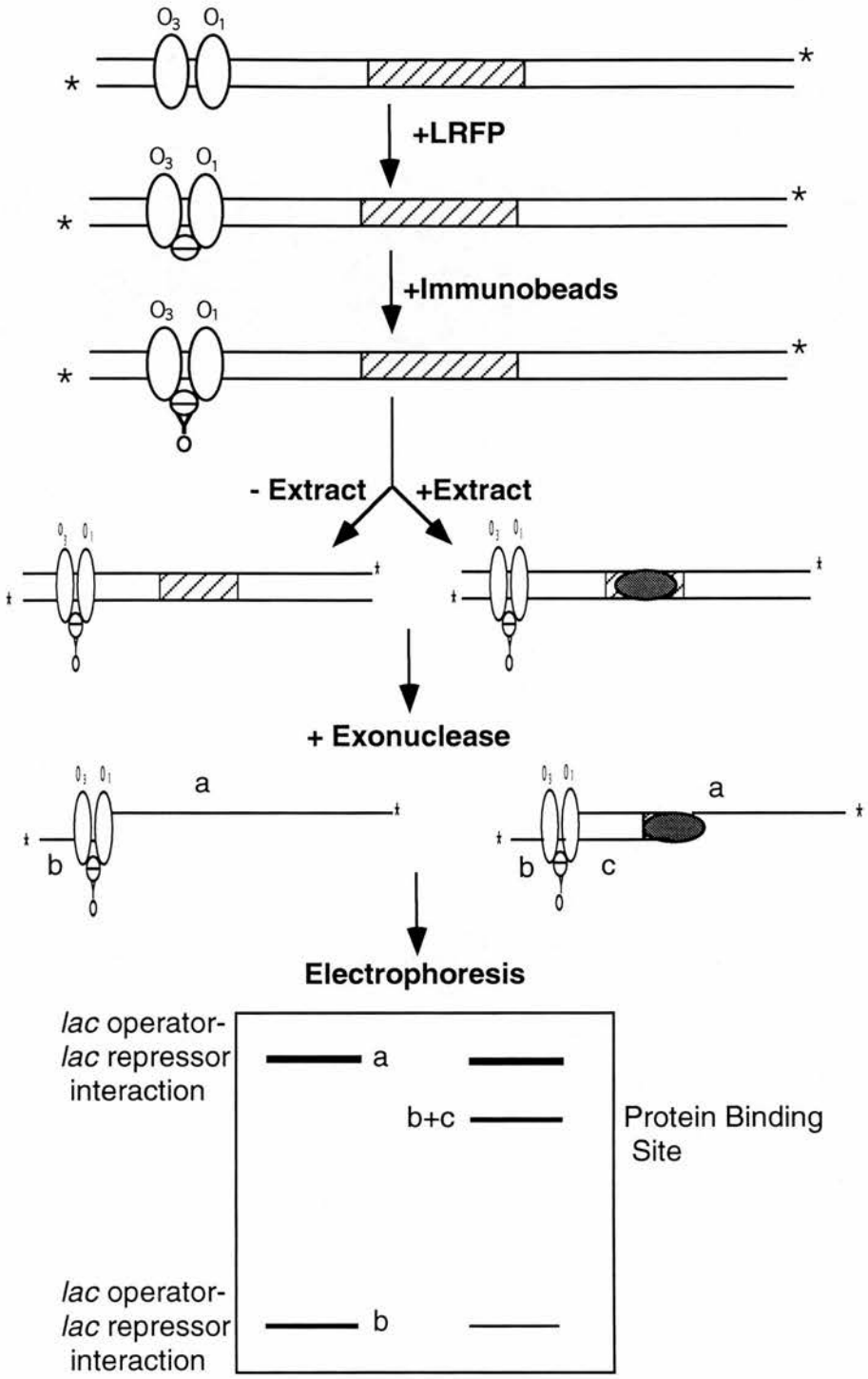
A *lac* repressor-β-galactosidase fusion protein (LRFP) () is added to the end-labelled DNA fragment allowing the *lac* repressor portion of the fusion protein to bind to the *lac* operator sequence.

Immunobeads carrying anti-β–galactosidase antibodies () are subsequently added. The anti-β–galactosidase antibody recognises the β-galactosidase portion of the fusion protein. This produces a complex of DNA linked to immunobeads via the *lac* repressor-β-galactosidase fusion protein.

The complex is then incubated with protein extract in the presence of the non-specific DNA competitor poly(d[I-C]), in order to allow sequence specific DNA-binding protein(s) () to bind to any recognition sites within the DNA.

T7 gene 6 exonuclease is then added, this enzyme digests the DNA in a 5' to 3' direction. In the absence of specific binding factor(s), exonuclease enzyme proceeds digestion along the DNA only stopping at the *lac* repressor-*lac* operator interaction. This produces two DNA fragments on electrophoresis corresponding to the *lac* operator-*lac* repressor interactions, one large (indicated as a) and one small (indicated as b).

If protein(s) are specifically bound to any point on the DNA fragment then they prevent exonuclease digestion proceeding along the DNA. This results in a diminution of the lower *lac* repressor-*lac* operator interaction (b) and the appearance of new band(s) (indicated as b + c) corresponding to the bound factors (Protein Binding Site).


O3 O1
+LRFP
+Immunobeads
- Extract
+Extract
+ Exonuclease
a
b
c
Electrophoresis
lac operator-lac repressor interaction
a
b+c
Protein Binding Site
lac operator-lac repressor interaction
b

Plasmid construction.

In order to analyse a region of the rPPT-A promoter close to the transcriptional start site and TATA box for potential regulatory elements the constructs pSM1/Q5 and pSM1'/Q5 were made. These constructs are derived from the plasmid pPPT (Quinn, 1992) which contains a fragment of the rPPT-A gene, spanning nucleotides -3356 to +447 relative to the major transcriptional start site, cloned into the polylinker of the vector pUC19.

Construction of pSM1/Q5 and pSM1'/Q5 was carried out as follows. Plasmid pPPT was digested in the polylinker with the restriction enzymes *Xba* I and *Eco*R I and a fragment of approximately 3900 bp, corresponding to rPPT-A sequence, was purified from a 1% low melting point agarose gel. Purified insert DNA was then digested with the restriction enzyme *Hinf* I and the desired fragment, covering nucleotides -160 to +447 of the rPPT-A promoter, was again purified from a low melting point agarose gel. Sticky ends were filled in with Klenow DNA polymerase.

The vector pUC19 was digested with *Sma* I and then dephosphorylated with CIP. Dephosphorylated vector and filled-in insert were ligated and transformed into competent HB101 cells. Resulting colonies were picked and tested for inserts. Clones containing inserts in both directions were grown-up and the sequence of the insert was confirmed by DNA sequencing. The constructs are termed pSM1/Q5 and pSM1'/Q5, and are diagramatically represented in Figure 6.

3.5.2. Identification of a protein/DNA interaction within the rPPT-A promoter by the exonuclease protection assay.

Exonuclease protection analysis was carried out using the construct pSM1/Q5 (Figure 6) and HeLa nuclear extract.

Figure 7 shows the result of such an assay. The *lac* operator-*lac* repressor interaction bands (indicated as *lac* operator interaction) are, as expected, only present when exonuclease enzyme was added (lanes 3 & 4). An additional strong band (indicated as Protein/DNA interaction) is present in lane 4 when both exonuclease enzyme and HeLa nuclear extract were included in the reaction. This band corresponds to a protein/DNA interaction site which, from the position of size marker DNA, lies approximately 300 bp from the end of the DNA fragment. Taking the pUC vector and rPPT-A DNA into account, the binding site of this factor can be placed close to the major transcriptional start site.

Thus, using the exonuclease protection assay system, a binding site for proteins expressed in HeLa cells has been identified close to the major transcriptional start site of the rPPT-A gene.

Figure 6. Diagrammatic representation of the constructs used for exonuclease protection and DNase 1 footprinting assays.

A region spanning nucleotides -856 to +447 of the rPPT-A promoter (indicated as rPPT-A) was analysed for DNA/protein interaction sites.

In order to achieve this a series of constructs were used. Constructs prPPT-βGal 3, 4 and 6 and pVL24 contain various fragments of the rPPT-A promoter linked to a *lac*Z (β-galactosidase) reporter gene (Supplied by K. Chapman, Department of Medicine, Western General Hospital, Edinburgh; Mulderry *et al.*, 1993). These are;

1. prPPT-βGal 3: A 957 bp fragment spanning nucleotides -865 to +92.
2. prPPT-βGal 4: A 763 bp fragment spanning nucleotides -671 to +92.
3. pVL29: A 523 bp fragment spanning nucleotides -431 to +92.
4. prPPT-βGal 6: A 437 bp fragment spanning nucleotides -345 to +92.

pSM1/Q5 and pSM1'/Q5 contain a 607 bp fragment spanning nucleotides -160 to +447 cloned into the pUC19 vector in both orientations. pSM1'/Q5/*Dde* I is derived from pSM1'/Q5.

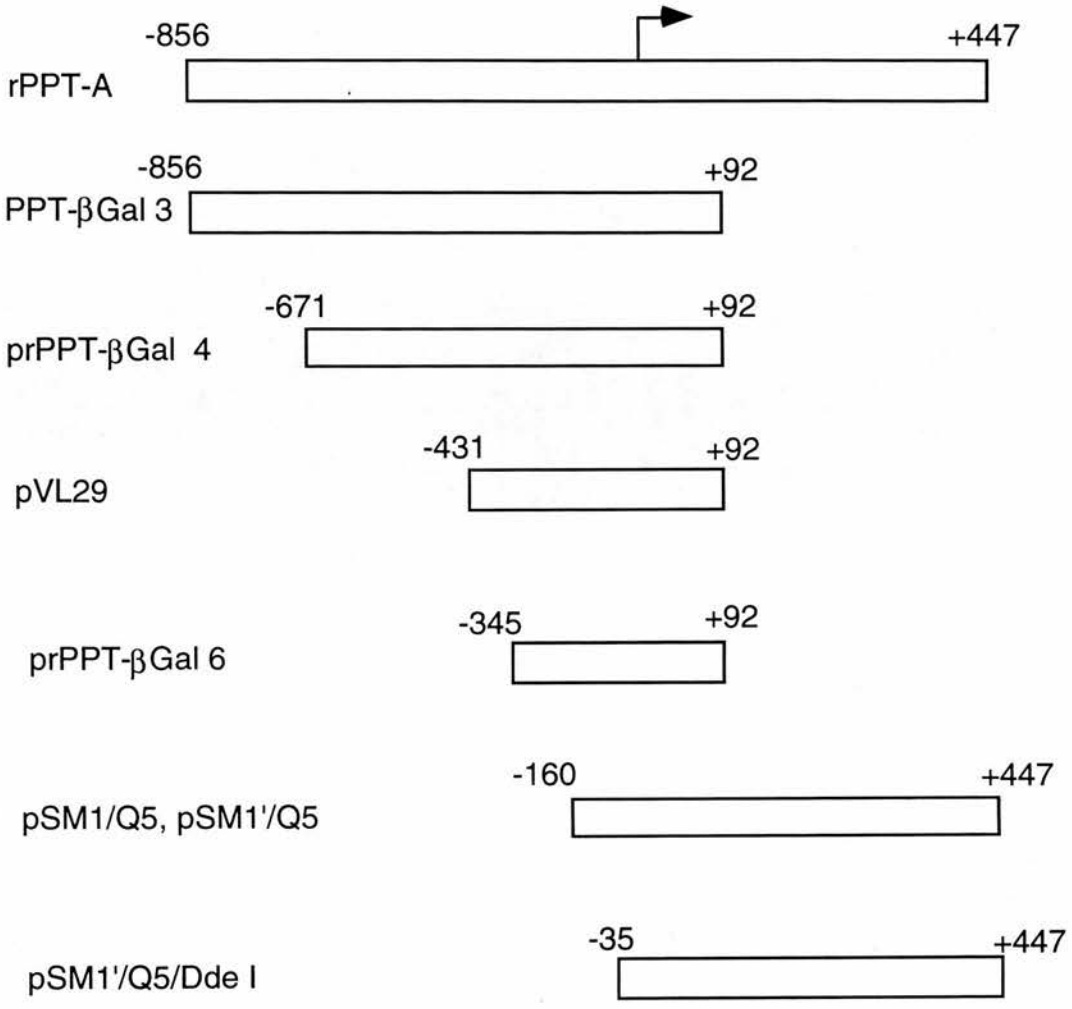
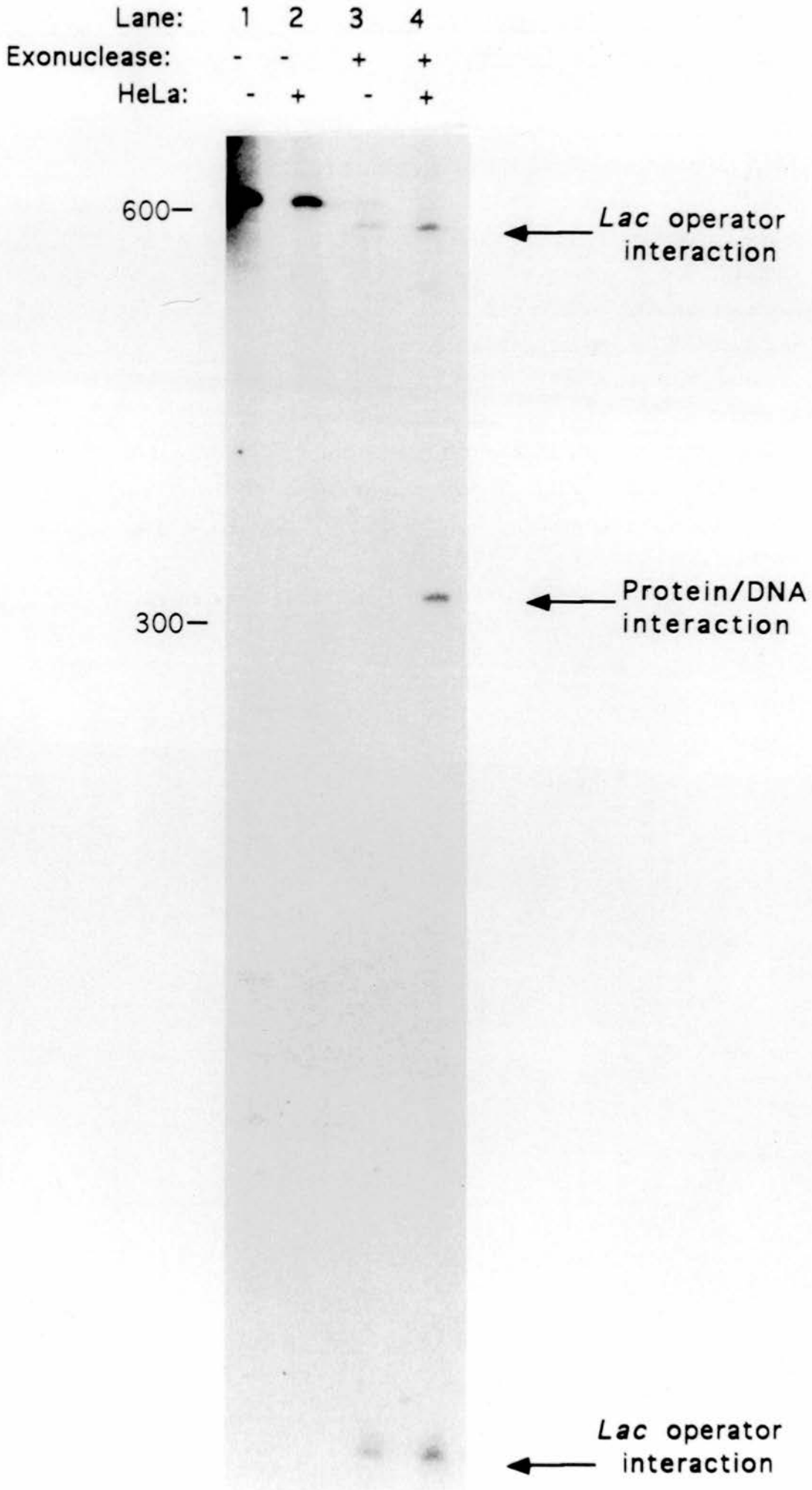

-856
+447
rPPT-A
-856
+92
PPT-βGal 3
-671
+92
prPPT-βGal 4
-431
+92
pVL29
-345
+92
prPPT-βGal 6
-160
+447
pSM1/Q5, pSM1'/Q5
-35
+447
pSM1'/Q5/Dde I

Figure 7. Identification of a protein/DNA interaction site within the rPPT-A promoter by the exonuclease protection assay.

The construct pSM1/Q5 was digested with the restriction enzyme *Pvu* II and a resulting DNA fragment, containing rPPT-A sequence and the *lac* operator sequence from pUC19, was radioactively 3' end labelled. 100ng of the end-labelled DNA fragment was digested with T7 gene 6 exonuclease, following incubation with HeLa nuclear extract.

Lane 1 shows the position of the labelled DNA fragment which, as indicated, co-migrates with size marker DNA at approximately 600 bp. Lane 2 demonstrates that the addition of HeLa nuclear extract (0.2mg) had no effect on the migration of the DNA fragment. The addition of exonuclease enzyme, (40 units) (lane 3), digested the DNA to reveal two *lac* operator-*lac* repressor interactions (indicated as *Lac* operator interaction). In the presence of HeLa nuclear extract, (0.2mg) (lane 4), exonuclease enzyme (40 units) digested the DNA to reveal an additional band which corresponds to a protein/DNA interaction site (indicated as Protein/DNA interaction). The approximate position of this interaction, by alignment with size marker DNA, is 300 bp from the end of the DNA fragment


Lane: 1 2 3 4
Exonuclease: - - + +
HeLa: - + - +
600
300
Lac operator interaction
Protein/DNA interaction
Lac operator interaction

Further exonuclease protection assays using the same construct (pSM1/Q5) with different extracts and conditions have demonstrated the presence of several more specific protein/DNA interaction sites, placed both 5' and 3' of the major transcriptional start site (personal communication, J.Quinn). Additionally, a previous study using the same assay identified multiple protein binding sites within a 3.8 kb fragment of the promoter (Quinn, 1992).

Unfortunately, the exonuclease protection assay does not allow the exact location of identified protein/DNA interaction sites to be determined. In order to achieve such accurate mapping of such sites and to examine them in more detail, DNase 1 footprinting and electrophoretic mobility shift assays were subsequently carried out, as described in the following sections.

3.6. Analysis of the rPPT-A promoter by DNase 1 footprinting and electrophoretic mobility shift assays.

DNase 1 footprinting analysis (Section 2.2.6.2.) uses singly 5' end labelled DNA fragments in order to identify sequence specific protein binding sites within a particular region of DNA sequence. This particular assay, although requiring the use of more protein extract than exonuclease protection assays, does allow the exact bases protected by bound factors to be accurately determined.

A series of constructs, diagramatically represented in Figure 6, were used in this study for DNase 1 footprinting assays. The construction of the plasmids pSM1/Q5 and pSM1'/Q5 is described in Section 3.5.1. Constructs prPPT-βGal 3, 4 and 6 and pVL24 (a gift from K. Chapman) contain various fragments of the rPPT-A promoter linked to a *lac*Z (β-galactosidase) reporter gene. These constructs were sequenced (Section 2.2.7.1.) and were shown to contain the following regions from the rPPT-A promoter:

1. prPPT-βGal 3; A 957 bp fragment spanning nucleotides -865 to +92.
2. prPPT-βGal 4; A 763 bp fragment spanning nucleotides -671 to +92.
3. pVL29; A 523 bp fragment spanning nucleotides -431 to +92.
4. prPPT-βGal 6; A 437 bp fragment spanning nucleotides -345 to +92.

In order to generate suitable substrates for use in DNase 1 footprinting assays constructs were linearised by restriction enzyme digestion in the polylinker with *Sal* I and then radioactively 5' end labelled (Figure 3). Each DNA fragment was then digested in the polylinker, with *Xba*1 for prPPT-βGal-3, 4 and 6 and pVL29 and with *Hin*d III for pSM1/Q5 and pSM1'/Q5. This produced two singly end labelled fragments, one of almost full length and the other of only 10-20 bp in length which would not interfere with the assay. Singly end labelled fragments were incubated with protein extract (approximately 1.0 mg) in the presence of the non-specific competitor

DNA, poly(d[I-C]). This allowed sequence specific proteins to bind to any DNA recognition sites present within the rPPT-A promoter sequence. DNase 1 enzyme was then added at a predetermined concentration in order to introduce, on average, a single random nick per DNA molecule. The reactions were terminated by the addition of EDTA and SDS and reaction products were visualised by denaturing polyacrylamide gel electrophoresis and autoradiography. Regions of DNA bound by sequence specific proteins were identified by comparing the ladder of DNase 1 digestion products in the presence of protein with that generated in the absence of protein. Bands missing from the ladder resulted from bound proteins protecting the DNA from DNase 1 nicking and darker bands, or hypersensitive sites, appeared when bound proteins altered the conformation of the DNA to allow preferential DNase 1 nicking. The exact bases protected by bound factors were determined by Maxam Gilbert sequence reactions of the appropriately labelled DNA fragment (Section 2.2.7.2.) electrophoresed along with the DNase 1 reaction products.

Electrophoretic mobility shift assays (Section 2.2.6.1.) (Singh *et al.*, 1986) were also used, in some cases, to analyse footprinted sequences from the rPPT-A promoter in more detail. The electrophoretic mobility shift assay is a method used to identify and characterise protein/DNA binding activities. It involves incubating cellular extracts (nuclear or whole cell) with a short radioactively labelled oligonucleotide probe containing the consensus sequence and flanking regions for the transcription factor of interest. When reaction mixtures are run on a native polyacrylamide gel, DNA bound by protein is seperated from free DNA and appears as a retarded band, the specificity of which can be determined by the addition of competitor DNA.

3.6.1. Identification of potential regulatory elements within the rPPT-A promoter.

By the use of DNase 1 footprinting and electrophoretic mobility shift assays several potential regulatory sequence elements have been identified within the rPPT-A promoter. The sequence of each element has been underlined in Figure 8 and each element has been numbered 1 to 19.

3.6.1.2. Element 1.

DNase 1 footprinting analysis was carried out using the construct pSM1/Q5 (Figure 6) and HeLa nuclear extract. Figures 9 A and 9B, lane 1 show the pattern of DNase 1 digestion when no extract was present. When HeLa nuclear extract was added to the reaction (Figure 9A, lane 2 and Figure 9B, lanes 2 and 3) a footprint, as indicated, was clearly visible. From the Maxam and Gilbert sequence reaction in Figure 9A (lane 3), the footprint can be located to a 20 bp sequence element spanning nucleotides -67 to -47. This footprint is termed E-box 1 as it includes an E-box

Figure 8. Nucleotide sequence of 1300bp of the rPPT-A gene indicating potential regulatory elements.

The nucleotide sequence is numbered on the left hand side with the first nucleotide (+1) of the major transcriptional start site as an adenine base (shown in bold type).
Underlined sequences are regions identified as protein/DNA interaction sites by DNase 1 footprinting analysis and electrophoretic mobility shift assays. The numbers underneath underlined sequences correspond to the following elements:

1. Element containing an E-box motif (E-box 1).
2. Element containing an E-box motif (E-box 2).
3. Element displaying homology to both CRE/ATF and AP-1 consensus sequences ('CRE').
4. Element rich in dG and dA residues (Purine-Rich Element).
5. Element containing an E-box motif (E-box 3).
6. Element containing a six out of seven base pair match to an AP-1 consensus sequence (AP-1').
7. Element containing a perfect AP-1 consensus sequence (AP-1).
8. Element rich in dG nucleotides (5' G-Rich).
9. Unidentified Element 1 (U.D. 1).
10. Element rich in dA and dT nucleotides containing an octamer-like consensus sequence (5' AT-Rich).
11. Elements which bind both single-stranded and double-stranded DNA binding proteins (dsDBP/ssDBP).
12. Region located between the TATA box and transcriptional start site (Minimal Element).
13-16. Elements displaying tissue specificity in their protein binding activity.
17. Element rich in dG nucleotides (3' G-rich).
18. Unidentified Element 2 (U.D. 2).
19. Element rich in dT nucleotides (3' T-rich).

-865 CTGCAGAGCT CCAAAGGTAA GCATCCAGCC TTTCTAGTCC CCCAACAAGG

-815 CTAAAGGGGA GAGAGGCACA ATTATCCTCT TCCCACCCCT TCTGCCTTCA

-765 GGGTGTGCCT GGGAAGAAGC TGTAGGGGAA CAAAAGATGC CTTAGAATGG

11

-715 CTGATGGGTA AGTTCTACAT GAGAAAGGAG GTTTAAATTC CTCTTTCCCC

-665 TAAATGTAAA ACAAACCTGC CTTCATCCTC TGAAGCGGGA GACCGGAAAC

10

-615 ACTTTTGCAG TGCTAGAGAA ATGAGAATAT TCTGACTGAT TTGGTGGGGA

9

-565 GGGGGGTTGG GGGGGTGTGT TCCAGCCCTA GATATAACAC CTCATAAACC

8

-515 TTAAGACACA TAAAGTAGAA ATGAAAGGAA AACCCCGCTT GCTTCATCCC

-465 TCTGAAGTGC TTGCTGGTGT CTTAGTATTA TTCACAAGGT TTTGCTGCTC

-415 AAGTTATTTG GCTGTCCTCA AAGCGCAATA TTCCCTGATG CCTCTTGAGA

-365 GAAAAGTTCC CTAAGTCCGA AGCATGAGTC ACTTCGCTCA GTTTTGATGA

7 6

-315 GTAATCTCAG GTGTCACTGA ACCTTGTTCG GAAGAAGAGG GGAGGGGGGC

5 4

-265 GTCAGATTTG CAGACGGAAG AAAACAGGTC TCTCTGGATT GGATGGCGAG

-215 ACCTCGACTT CCCTAAAATT GCGTCATTTC GAACCCAATT TGGTCCAGAT

3 2

-165 GTTATGGACT CCGACGGGTT ACCGTCTCGG AAACTCTATC ACGCAAGCAA

-115 AAGGCGAGGG GGCGGCTAAT TAAATATTGA GCAGAAAGTC GCGTGGGGAG

-65 AGTGTCACGT GGCTCTCCAG GCTCATCACG CCTGAGATAA ATAAGGCGAA

1

-15 GCAGGAGCAG GGACTAGAGC GCACTCGGAC CAGCTCCACT CCAGCACCGC

12

+36 GGCGGAGGAG AGCGAGGAGC GCCCAGCAAG TGCGCACCTG CGGAGCATCA

13 14

+86 CCGGGTCCGA CCGCAGTGAG TACCACTCCC GCCTGCCATC TGCCGCTGCC

15 16

+136 TGGGGCCCGT CGTGCCTTGG TCAGAGGTGG CGTCCGCCCG GGGTTTCACC

+186 TGCAGCAGCC ACAGACCCGG AGAGAAACTG TGCAAATACC AACACCTCTC

+236 TTTGTCTGCC TTCAGGCTTC GGAGTGTGGG TCAGTGGGTA GGGGGCTGGG

17

+286 ACGTTGAGAG GCAAAGAGAG GAGGACTTGA GGCTCTTTGG CACGTCAGTA

18

+336 GCCTTCTTAA AGGGTTTGGA GAAATGTTCG ACCTGGTGGG TTTTTGTTTT

+386 TGTTTTAATC TTGGTTTAGC GAGACCTCTT CTTTCTTCCT TGTGGCATTG

19

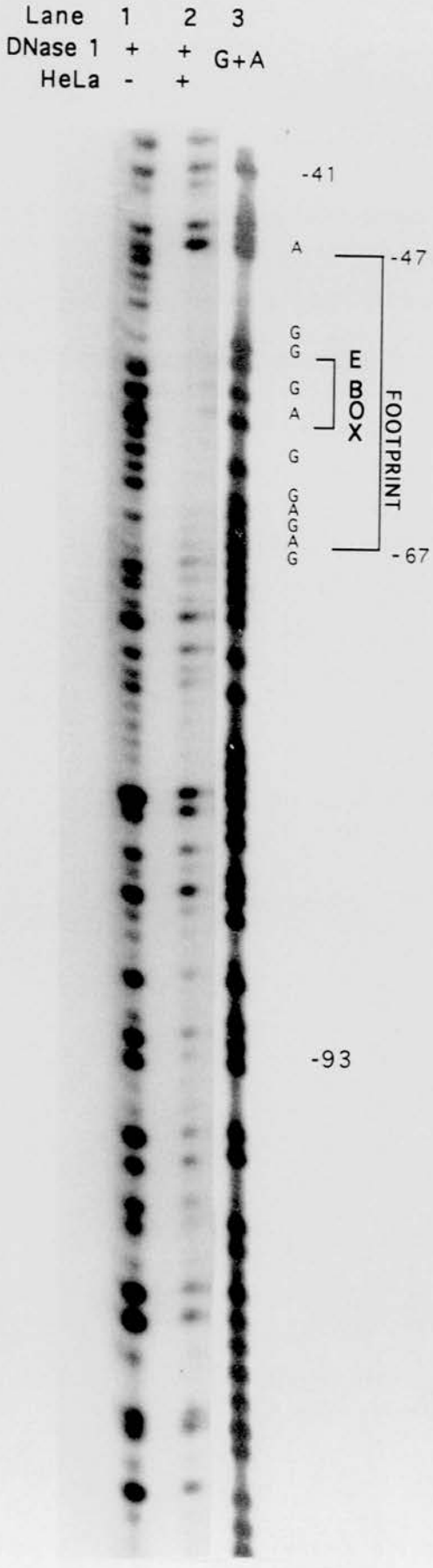
+436 AAATTCAGCT GC

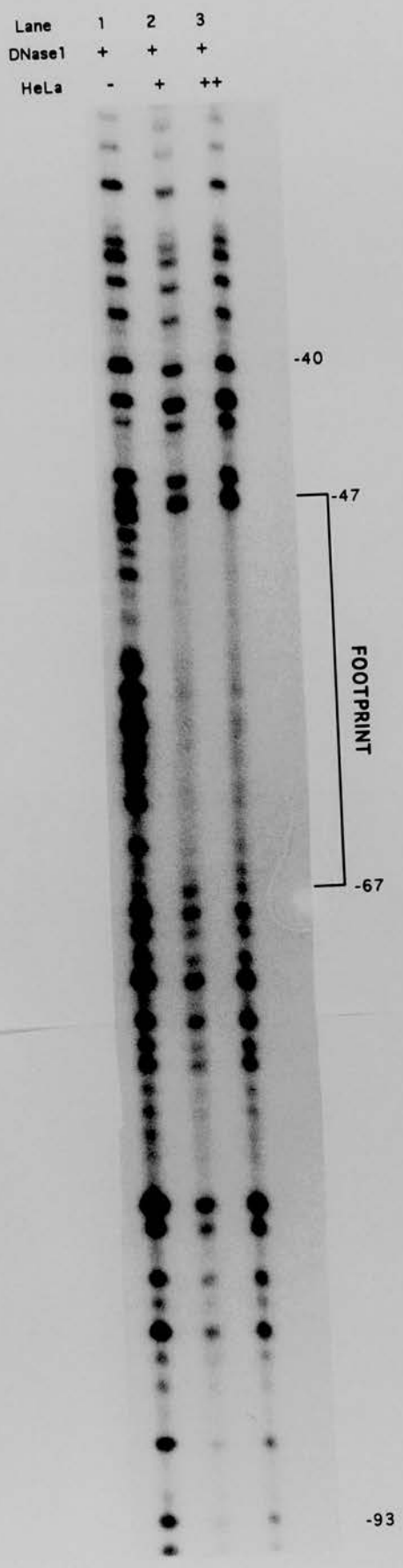
Figure 9. DNase 1 footprinting analysis of a region of the rPPT-A promoter encompassing an E-box motif.

The construct pSM1/Q5 was singly 5' end labelled with T4 polynucleotide kinase and 100ng of labelled probe was digested with DNase 1, in the absence or presence of HeLa nuclear extract.

A. Lane 1 shows the pattern of digestion generated by 1.5 units of DNase 1. In addition to DNase 1 (30 units), in lane 2 HeLa nuclear extract (1mg) was added and lane 3 shows the G+A Maxam Gilbert sequencing reaction of the DNA probe.
A footprint was generated (lane 2) which, from the sequence in lane 3, spans nucleotides -67 to -47 relative to the major transcriptional start site. This element is termed E-box 1 as it contains an E-box motif (CACGTG).

B. Lane 1 shows the pattern of digestion generated by 1.5 units of DNase 1. In addition to DNase 1 (30 units), increasing amounts of HeLa nuclear extract were added in lanes 2 and 3 (0.5 and 1mg respectively).
A footprint was generated (lanes 2 and 3) which, from comparison with 9A, spans nucleotides -67 to -47 relative to the major transcriptional start site.


Lane 1 2 3
DNase 1 + + G+A
HeLa - +
-41
A
-47
G
G
G
A
G
G
A
G
A
G
E BOX
FOOTPRINT
-67
-93


Lane 1 2 3
DNase1 + + +
HeLa - + ++
-40
-47
FOOTPRINT
-67
-93

consensus sequence (CANNTG) (shown in bold type below), recognised by members of the bHLH and bHLH/ZIP proteins (Section 1.2.3.1.1.h.):

-67 AGAGTGT**CACGTG**GCTCTCCA -47

In order to further examine the interactions between the E-box 1 element and proteins from various cell lines and tissues an oligonucleotide, termed E1 (Figure 10), was made containing sequence spanning nucleotides -69 to -48 from the rPPT-A promoter, for use in electrophoretic mobility shift assays.

Figure 11 shows the result of such an electrophoretic mobility shift assay. HeLa nuclear extract (lanes 1-3) and C1300 (lanes 4-6), PC12 (lanes 7-9) and AtT20 (lanes 10-12) whole cell extracts all formed complexes with the probe (indicated by an arrow), with similar mobilities (compare lanes 1, 4, 7, and 10). These complexes were demonstrated to be sequence specific as they were competed by specific DNA competitor (lanes 2, 5, 8, and 11) and not by a heterologous competitor, (PPT AP-1, Figure 10) (lanes 3, 6, 9, and 12). Non-specific complexes formed between the E1 oligonucleotide and the various extracts are indicated as N.S.

Figure 12 shows that sequence specific complexes (indicated by an arrow) were also formed between oligonucleotide E1 and proteins from rat spinal cord and caudate nuclear extracts. These complexes display a similar mobility to the complex formed with HeLa nuclear extract (compare lanes 1, 2 and 5).

3.6.1.3. Element 2.

DNase 1 footprinting assays carried out using the construct prPPT-βGal6 (Figure 6) and HeLa nuclear extract revealed a footprinted region spanning nucleotides -177 to -155 (Figure 13, lane 2).

This footprint contains an E-box consensus sequence (CANNTG), termed E-box 2, as shown in bold type below;

-177 TTTGGTC**CAGATG**TTATGGACTC -155

Electrophoretic mobility shift assays, using this footprinted sequence as an oligonucleotide probe, have demonstrated that the binding of proteins present in HeLa cells to this element is specifically competed by related E-box containing oligonucleotides (personal communication, J. Quinn).

3.6.1.4. Element 3.

DNase 1 footprinting assays using the construct prPPT-βGal6 (Figure 6) and HeLa nuclear extract also revealed a footprint spanning nucleotides -198 to -180 (Figure 13, lane 2), as shown below;

-198 AT**TGCGTCA**TTTCGAACCC -180

Figure 10. Sequence of the oligonucleotides used in electrophoretic mobility shift assays.

The sequence of the oligonucleotides used in electrophoretic mobility shift assays are shown in their double stranded forms.

Oligonucleotides E1, PPT AP-1, PPT 'CRE', 3' G-Rich, 5' AT-Rich, 3' T-Rich and 168/169 contain the following sequences from the rPPT-A promoter, relative to the major transcriptional start site:

E1; -69 to -48.

PPT AP-1; -347 to -327.

PPT 'CRE'; -200 to -186.

3' G-Rich; +264 to +291.

5' AT-Rich (782/783); -672 to -651.

3' T-Rich; +369 to +395.

168/169; -20 to +4

POB contains an element form the pro-opiomelanocortin (POMC) gene spanning nucleotides -23 to +5 (Riegel *et al.*, 1990).

NS (Non-specific) is a randomly computer generated sequence.

Oct contains an octamer binding protein element from the immunoglobulin heavy-chain enhancer (Augereau and Chambon, 1986; and Staudt *et al.*, 1988).

E1	5' TCGAGGAGAGTGTCACGTGGCTCTCC CCTCTCACAGTGCACCGAGAGGAGCT
PPT AP-1	5' TCGAAAGCATGAGTCACTTCGCTC TCGTACTCAGTGAAGCGAGAGCT
PPT 'CRE'	5' TCGAAATTGCGTCATTTC TTAACGCAGTAAAGAGCT
3' G-Rich	5' TCGAGGTCAGTGGGTAGGGGGCTGGGACGTTG CCAGTCACCCATCCCCCGACCCTGCAACAGCT
5' AT-Rich (782/783)	5' TCGATTTCCCCTAAATGTAAAACAAA AAAGGGGATTTACATTTTGTTTAGCT
3' T-Rich	5' TCGATGGTGGGTTTTTGTTTTTGTTTTAATC ACCACCCAAAAACAAAAACAAAATTAGAGCT
168/169	5' TCGAGCGAAGCAGGAGCAGGGACTAGAG CGCTTCGTCCTCGTCCCTGATCTCAGCT
POB	5' AAGAAGAGAGAAGAGTGACAGGGACCAA TTCTTCTCTCTTCTCACTGTCCCTGGTT
NS	5' ATCCCTTTAAATTTGCGAGCT GGAAATTTAAACGCTCGACTA
Oct	5' GAGATCTAGCATGCAAATCATTGT TAGATCGTACGTTTAGTAACAGAG

Figure 11. Electrophoretic mobility shift assay using an E-box motif from the rPPT-A promoter and extracts from HeLa, C1300, PC12 and AtT20 cells.

A double stranded oligonucleotide (E1) from the rPPT-A promoter spanning nucleotides -69 to -48, containing an E-box motif, was 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotide (1ng) was incubated with either HeLa nuclear extract or C1300, PC12 or AtT20 whole cell extracts (20µg each), in the presence of 1µg of poly d(I-C).

Lanes 1, 4, 7 and 10 demonstrate that several complexes were formed between the probe and HeLa, C1300, PC12 and AtT20 extracts respectively, in the absence of competitor DNA. One of these complexes, indicated by an arrow, was shown to be specific as it was competed by homologous oligonucleotide, (S) (100ng) (lanes 2, 5, 8 and 11), and not by a non-specific oligonucleotide, (NS) (100ng) (lanes 3, 6, 9 and 12). Non-specific complexes are indicated by N.S.

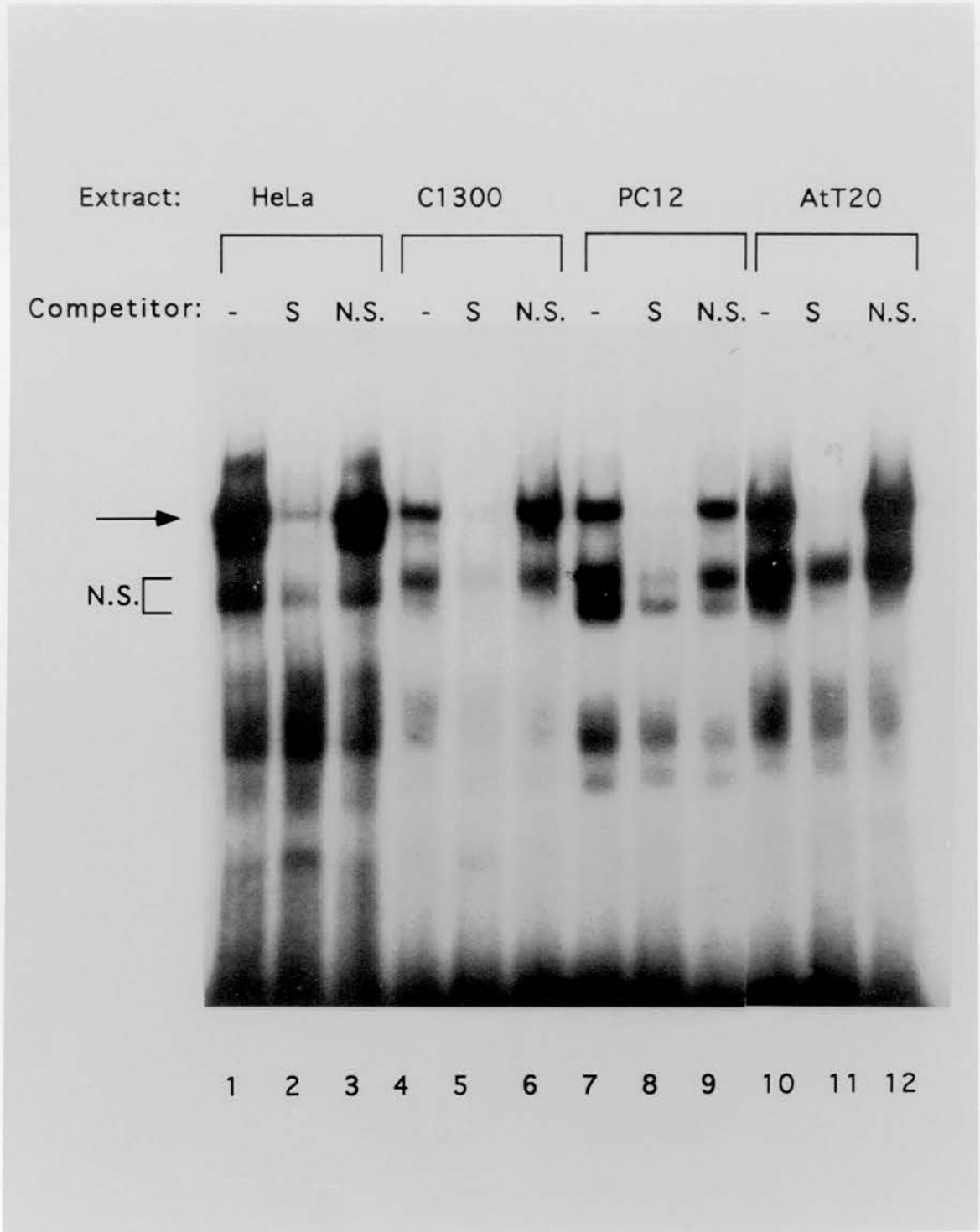

Extract:
HeLa
C1300
PC12
AtT20
Competitor:
- S N.S. - S N.S. - S N.S. - S N.S.
N.S.
1 2 3 4 5 6 7 8 9 10 11 12

Figure 12. Electrophoretic mobility shift assay using an E-box motif from the rPPT-A promoter and extracts from HeLa cells, rat spinal cord and rat caudate.

A double stranded oligonucleotide (E1) from the rPPT-A promoter spanning nucleotides -69 to -48, containing an E-box motif, was 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotide (1ng) was incubated with either HeLa nuclear extract, rat spinal cord or caudate nuclear extracts (20µg each), in the presence of 1µg of poly d(I-C).

Lanes 1, 2, and 5 demonstrate that complexes were formed between the probe and HeLa, rat spinal cord and rat caudate extracts respectively, in the absence of competitor DNA. One of the complexes, indicated by an arrow, formed between the rat spinal cord and caudate extracts was shown to be specific as it was competed by homologous oligonucleotide, (S) (100ng) (lanes 3 and 6), and not by a non-specific oligonucleotide, (NS) (100ng) (lanes 4 and 7).

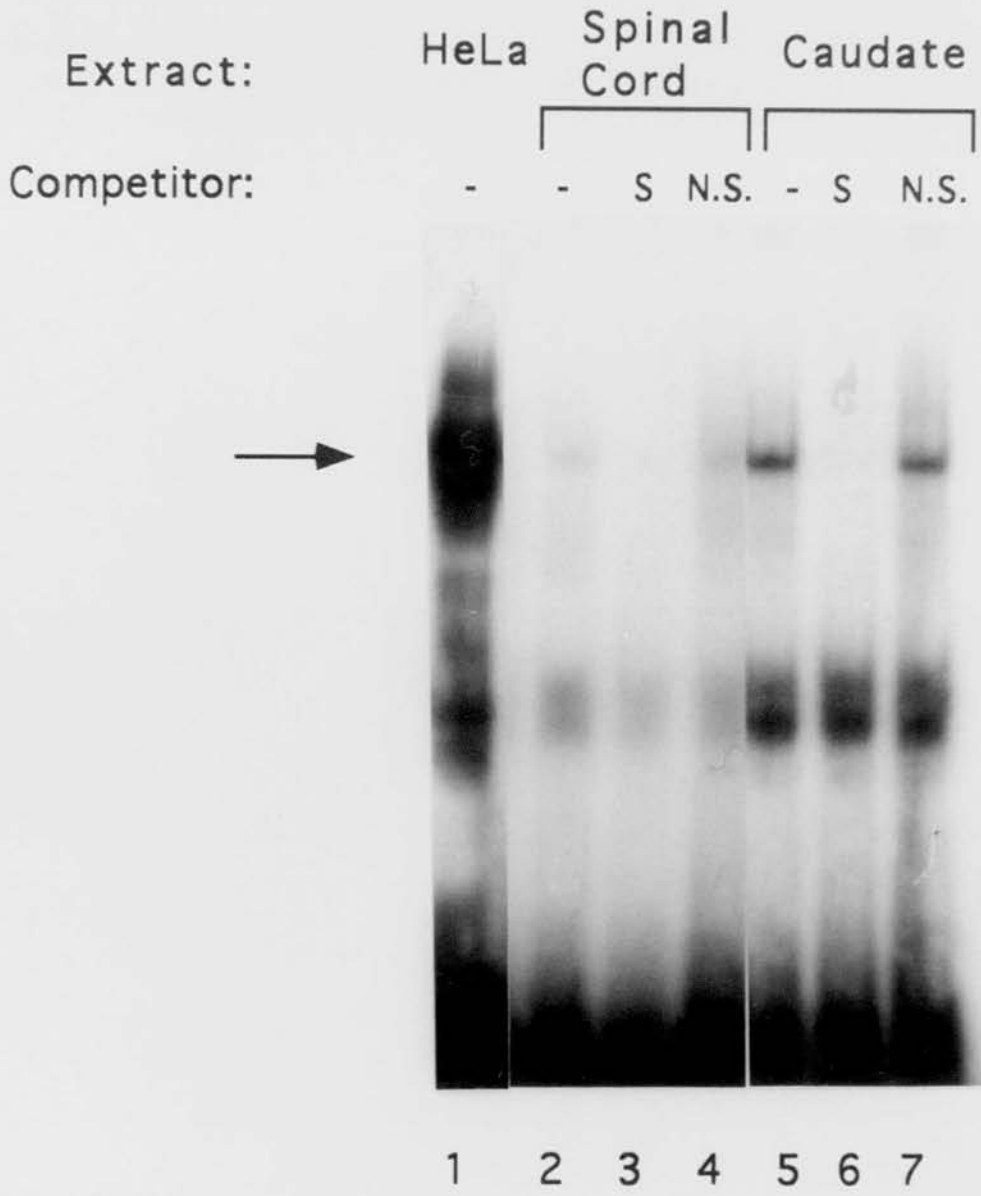

Extract:
HeLa
Spinal Cord
Caudate
Competitor:
-
-
S
N.S.
-
S
N.S.
1
2
3
4
5
6
7

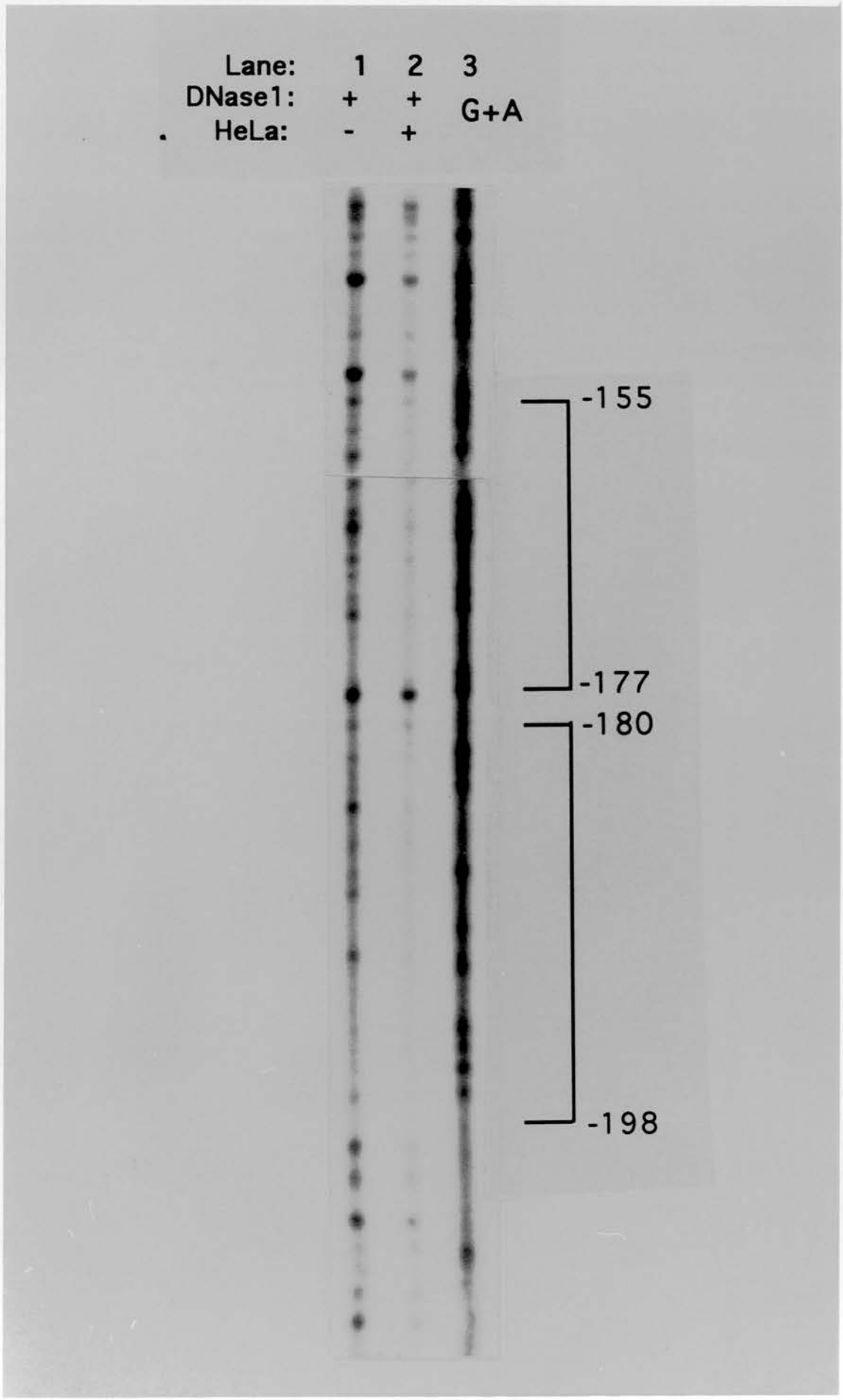
Figure 13. DNase 1 footprinting analysis of a region of the rPPT-A promoter encompassing a 'CRE' and an E-box motif.

The construct prPPT-βGal 6 was singly 5' end labelled with T4 polynucleotide kinase and 100ng of labelled probe was digested with DNase 1, in the absence or presence of HeLa nuclear extract.

Lane 1 shows the pattern of digestion generated by 1.25 units of DNase 1. In addition to DNase 1 enzyme (30 units), in lane 2 HeLa nuclear extract (1mg) was added. Lane 3 shows the G+A Maxam Gilbert sequencing reaction of the DNA probe.

Two footprints were generated in lane 2 which, from the sequence in lane 3, can be located to nucleotides -198 to -180 and -177 to -155.

The footprinted region spanning nucleotides -198 to -180 contains a sequence displaying homologies with both an AP-1 and a CRE/ATF consensus sequence and so is termed 'CRE'. The footprint spanning nucleotides -177 to -155 contains an E-box consensus sequence and so is termed E-box 2.


Lane: 1 2 3
DNase1: + + G+A
HeLa: - +
-155
-177
-180
-198

The region shown in bold type displays homology to both a CRE/ATF (cyclic AMP element binding protein/activating transcription factor) binding motif (TGACGTCA) and to an AP-1 (TGA G/C TCA) consensus sequence.

As this region of the promoter has previously been suggested to represent a functional CRE in the bovine PPT-A gene (Kageyama *et al.*, 1991) and has previously been noted as a potential CRE by others (Chapman *et al.*, 1993), it is referred to here as a 'CRE'.

3.6.1.5. Element 4.

A region containing a highly purine rich sequence was footprinted using the construct pVL29 (Figure 6) and HeLa nuclear extract, (Figure 14, lane 2). This element spans nucleotides -284 to -264 as shown below;

-284 AAGAAGAGGGGAGGGGGGCGT -264

It can be seen from Figure 14 that the footprint generated covering this region of the promoter is not as strong as those generated over the regions spanning nucleotides -324 to -308 and -345 to -330. However, an oligonucleotide containing the purine rich footprinted region and flanking sequence will form specific complexes in electrophoretic mobility shift assays, confirming that protein(s) will bind to this region of the promoter *in vitro* (personal communication, J.Quinn).

3.6.1.6. Element 5.

DNase 1 footprinting using the construct pVL29 (Figure 6) and HeLa nuclear extract failed to generate a footprint over a noted E-box motif within the rPPT-A promoter (Figure 14, lane 2). This element, shown in bold type below, is termed E-box 3:

-311 TCT**CAGGTG**TCA -300

However, an oligonucleotide spanning nucleotides -312 to -295, which includes the E-box 3 motif, will form a specific complex in electrophoretic mobility shift assays with HeLa nuclear extract (personal communication, J.Quinn). The complex formed has been shown by cross-competition analysis to be related to other complexes formed using E-box containing oligonucleotides. Thus, when in isolation the E-box 3 motif will bind sequence-specific proteins.

3.6.1.7. Element 6.

DNase 1 footprinting analysis using the construct pVL29 (Figure 6) and HeLa nuclear extract did, however, generate a footprint spanning nucleotides -324 to -308 (Figure 14, lane 2). The sequence of the footprinted region is shown below:

Figure 14. DNase 1 footprinting analysis of a region of the rPPT-A promoter encompassing two AP-1 motifs and a purine-rich element.

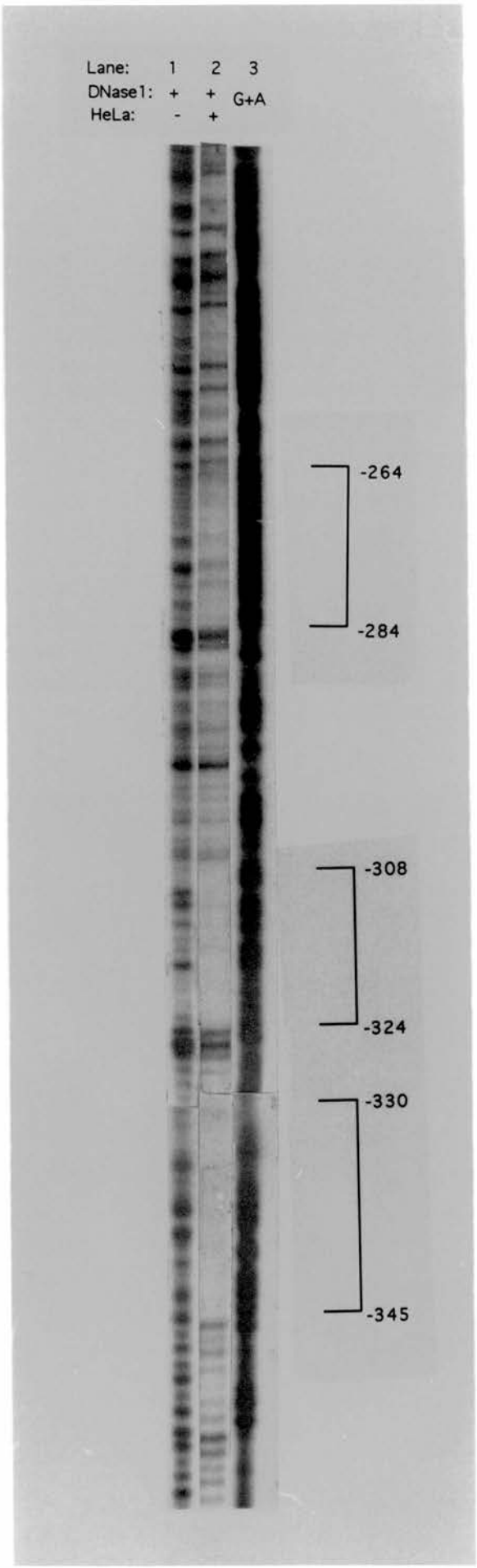
The construct pVL29 was singly 5' end labelled with T4 polynucleotide kinase and 100ng of labelled probe was digested with DNase 1, in the absence or presence of HeLa nuclear extract.

Lane 1 shows the pattern of digestion generated by 1.25 units of DNase 1. In addition to DNase 1 (30 units), in lane 2 HeLa nuclear extract (1mg) was added and lane 3 shows the G+A Maxam Gilbert sequencing reaction of the DNA probe.

Three footprints were generated in lane 2 which, from the sequence in lane 3, can be located to nucleotides -345 to -330, -324 to -308 and -284 to -264.

The footprint spanning nucleotides -345 to -330 contains a perfect AP-1 consensus sequence and is termed AP-1. The footprint spanning nucleotides -324 to -308 contains a six out of seven base pair match to an AP-1 consensus sequence and is termed AP-1' and finally the footprint spanning nucleotides -284 to -264 is rich in dG and dA nucleotides and is termed a purine-rich element.

Immediately downstream of the AP-1' footprint there is an E-box motif (CAGGTG, termed E-box 3) which is not footprinted.


Lane: 1 2 3
DNase1: + + G+A
HeLa: - +
-264
-284
-308
-324
-330
-345

-324 TTTTGA**TGAGTAA**TCTC -308

This sequence element contains a region (shown in bold type) which displays a six out of seven base pair match to an AP-1 consensus sequence (TGA G/C TCA) and so is termed AP-1'.

3.6.1.8. Element 7.

DNase 1 footprinting analysis using the construct pVL29 (Figure 6) and HeLa nuclear extract also generated a footprint spanning nucleotides -345 to -330 (Figure 14, lane 2), as shown below:

-345 AGCA**TGAGTCA**CTTCG -330

This footprint contains a perfect AP-1 consensus sequence (TGA G/C TCA) (shown in bold type) and so is termed AP-1.

3.6.1.9. Element 8.

DNase 1 footprinting analysis was carried out using the construct prPPT-βGal4 (Figure 6) and HeLa nuclear extract. A footprint, spanning nucleotides -575 to -546, was generated (Figure 15, lane 2), as shown below:

-575 TTGGTGGGGAGGGGGGTTGGGGGGGTGTGT -546

This footprint is rich in dG residues and so is termed a 5' G-rich element.

3.6.1.10. Element 9.

DNase 1 footprinting analysis carried out using the construct prPPT-βGal4 (Figure 6) and HeLa nuclear extract also generated a footprint spanning nucleotides -618 to -607 (Figure 15, lane 2), as shown below:

-618 AACACTTTTGCA -607

An oligonucleotide containing this element will form specific complexes with HeLa nuclear extract by electrophoretic mobility shift assays (personal communication, J. Quinn). As the identity of the proteins binding to the element have yet to be determined it has been termed unidentified element 1 (U.D. 1).

3.6.1.11. Element 10.

Inspection of 5' flanking sequence of the rPPT-A promoter reveals the presence of an AT-rich sequence, termed 5' AT-rich, which spans nucleotides -662 to -655 and, interestingly, displays a six out of eight base pair match with the octamer binding protein consensus sequence (ATGCAAAT), as shown below in bold type:

-665 TAA**ATGTAAAA**CAAA -651

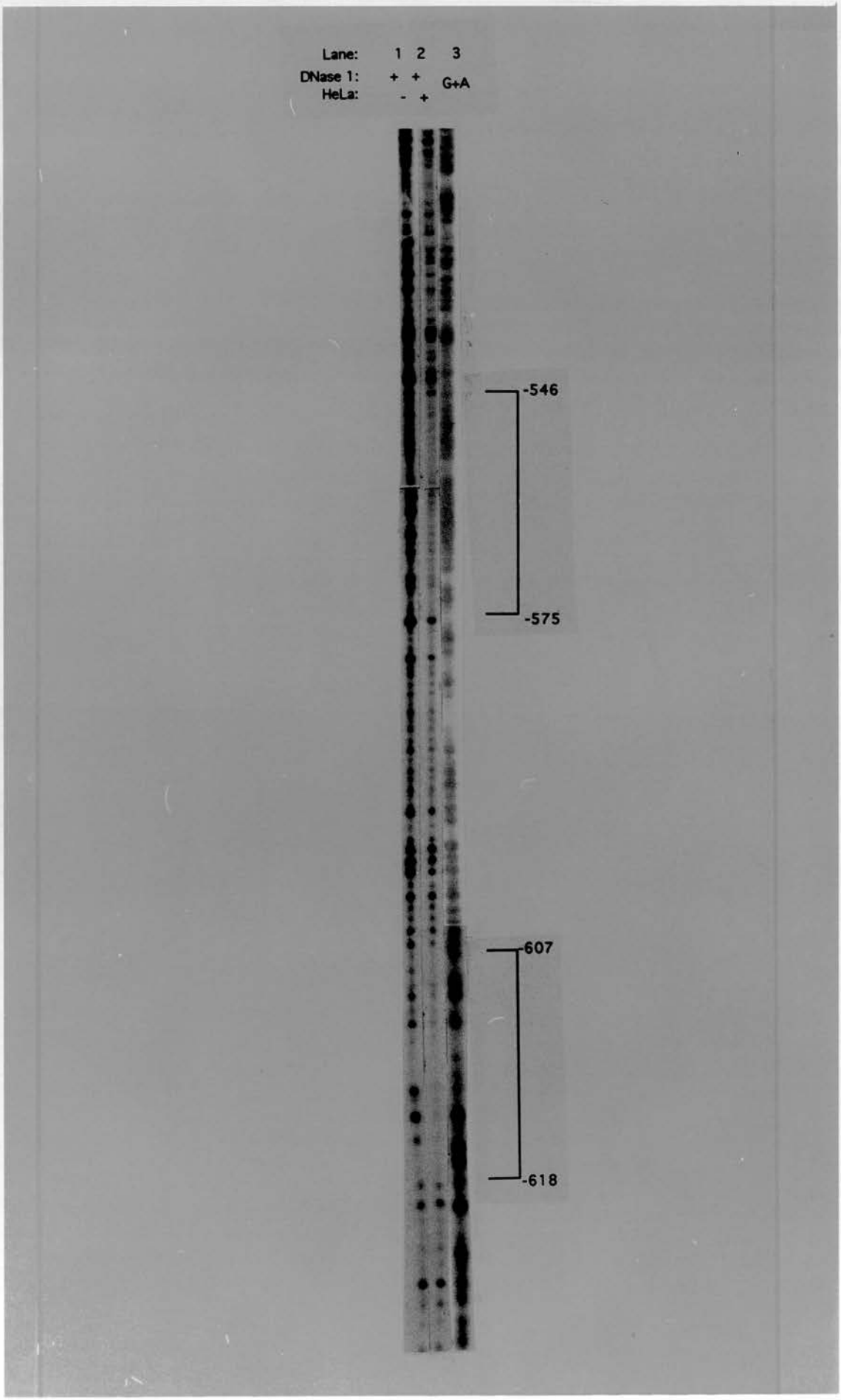
Figure 15. DNase 1 footprinting analysis of a region of the rPPT-A promoter encompassing U.D.1 and a G-rich sequence.

The construct prPPT-βGal 4 was singly 5' end labelled with T4 polynucleotide kinase and 100ng of labelled probe was digested with DNase 1, in the absence or presence of HeLa nuclear extract.

Lane 1 shows the pattern of digestion generated by 2.5 units of DNase 1. In addition to DNase 1, in lane 2 HeLa nuclear extract (1mg) was added and lane 3 shows the G+A Maxam Gilbert sequencing reaction of the DNA probe.

Two footprints were generated (lane 2) which, from the sequence reaction (lane 3), span nucleotides -618 to -607 and -575 to -546.

The footprint spanning nucleotides -618 to -607 is termed U.D. 1 and the footprint spanning nucleotides -575 to -546 is rich in dG nucleotides and is termed 5' G-Rich.


Lane: 1 2 3
DNase 1: + + G+A
HeLa: - +
-546
-575
-607
-618

Footprinting analysis of this region of the promoter using the construct prPPT-βGal3 (Figure 6) and HeLa nuclear extract failed to detect any protein binding activity over the 5' AT-rich sequence. However, closer inspection of such footprinting assays reveals that this region of the promoter is not cleaved by DNase 1, even in the absence of protein. Therefore if a protein does bind over this region a footprint will not be apparent.

The fact that 5' AT-rich sequence displays a high degree of homology with the octamer binding protein consensus sequence suggests that proteins belonging to the octamer binding protein family may recognise and specifically bind to this region of the rPPT-A promoter. In order to determine if this is so an oligonucleotide was made corresponding to the 5' AT-rich element and flanking sequences, spanning nucleotides -672 to -651 (Figure 10, 5' AT rich), for use in electrophoretic mobility shift assays.

Figure 16, lane 1 shows that a complex, indicated by an arrow, was formed between the 5' AT rich oligonucleotide (termed 782/783) and HeLa nuclear extract. This complex was shown to be specific as it was competed by increasing concentrations of homologous DNA competitor, (Specific) (lanes 2-4), and not by increasing concentrations of a non-specific competitor DNA, (NS, Figure 10) (lanes 11-13). Increasing concentrations of an oligonucleotide containing the octamer binding element from the immunoglobulin heavy-chain enhancer, (Octamer) (Augereau and Chambon, 1986; Staudt *et al.*, 1988) (Oct, Figure 10), also competed for complex formation (lanes 5-7). In fact competition was complete with only a 10-fold concentration excess of competitor (lane 5). Lane 14 shows the complex formed between the ubiquitous Oct-1 protein from HeLa cells and the octamer containing oligonucleotide. Comparison of lane 1 with lane 14 shows that the Oct-1 complex migrates to a similar position as that formed with the 5' AT-rich oligonucleotide. Therefore, it appears that the 5' AT-rich element from the rPPT-A promoter will bind octamer binding proteins present in HeLa cells.

Electrophoretic mobility shift assays were also carried out using the octamer binding element containing oligonucleotide (Oct) as a probe. Figure 17, lane 1 shows the position of the complex, indicated by an arrow, formed between the octamer oligonucleotide and Oct-1 from HeLa nuclear extract. The formation of this complex was competed by increasing concentrations of both homologous oligonucleotide, (Specific) (lanes 2-4), and 5' AT-rich oligonucleotide, (5') (lanes 5-7), but not by a non-specific oligonucleotide, (lanes 11 and 12). In Figure 18 it is shown that the 5' AT-rich oligonucleotide will also compete for complex formation between the octamer oligonucleotide and proteins present in C1300 whole cell extract. Lane 1 shows the position of the complex, indicated by an arrow, formed between Oct-1 from HeLa nuclear extract and the octamer containing oligonucleotide. This complex migrates in a

Figure 16. Electrophoretic mobility shift assay using the rPPT-A 5' AT-rich element and extract from HeLa cells.

A double stranded oligonucleotide (termed 782/783) containing the rPPT-A 5' AT-rich element, spanning nucleotides -672 to -651 relative to the major transcriptional start site, was 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotide (1ng) was incubated with HeLa nuclear extract (20μg), in the presence of 1μg of poly d(I-C).

Several complexes were formed between the probe and HeLa cell extract, in the absence of competitor DNA (lane 1). One of these complexes, indicated as an arrow, was shown to be specific as it was competed by increasing concentrations (10, 100 and 200ng) of homologous oligonucleotide (Specific) (lanes 2-4). Complex formation was also competed by increasing concentrations (10, 100 and 200ng) of oligonucleotides containing either the immunoglobulin octamer binding element, (Octamer) (lanes 5-7), or a 3' T-rich sequence from the rPPT-A promoter (lanes 8-10). Increasing concentrations (10, 100 and 200ng) of non-specific oligonucleotide (lanes 11-13) failed to compete for complex formation.

Lane 14 shows a specific complex (indicated as an arrow) formed between the immunoglobulin octamer binding element containing oligonucleotide (Octamer) and HeLa nuclear extract.

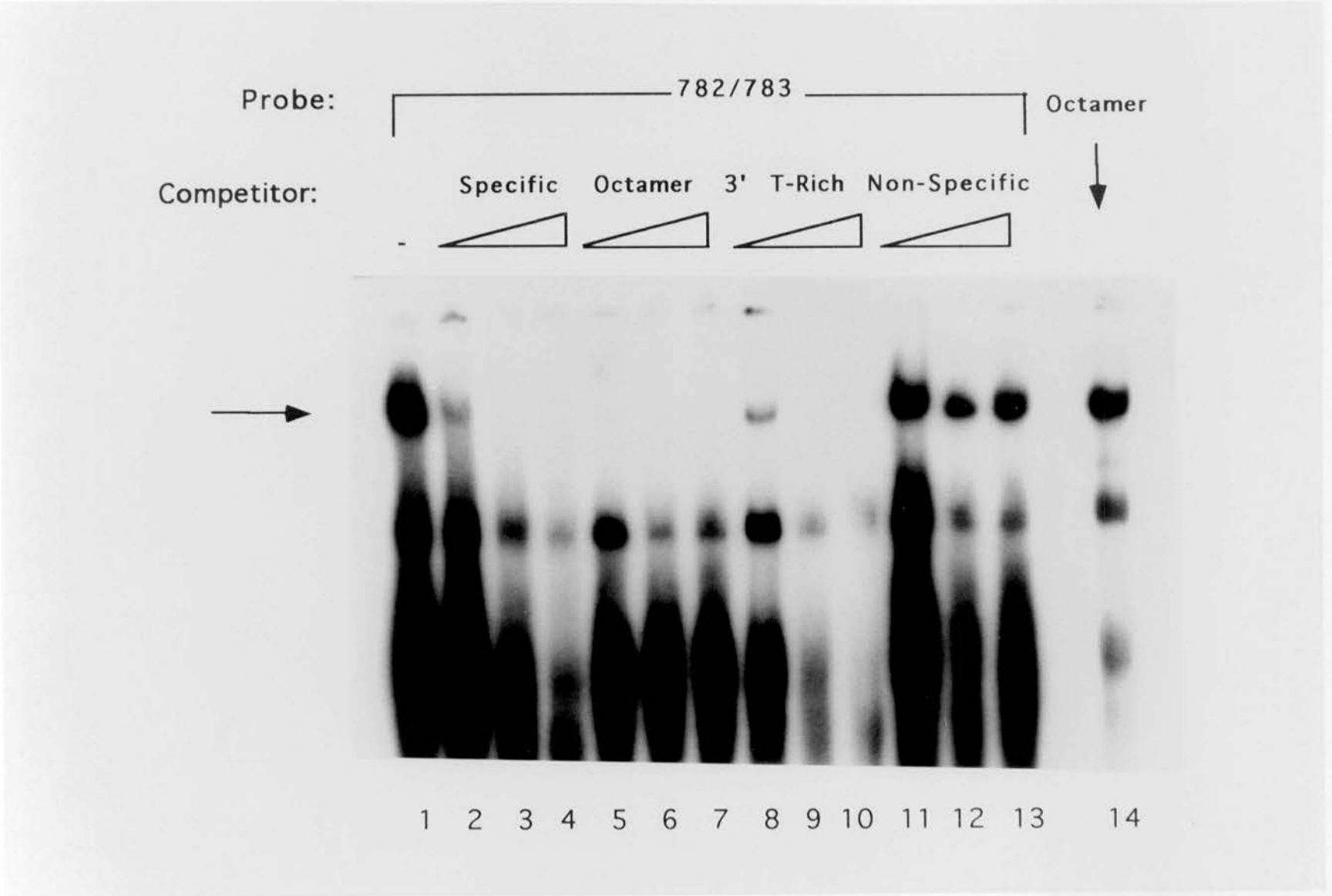

Probe:
782/783
Octamer
Competitor:
Specific
Octamer
3' T-Rich
Non-Specific
-
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Figure 17. Electrophoretic mobility shift assay using the immunoglobulin octamer binding element and extract from HeLa cells.

A double stranded oligonucleotide containing the immunoglobulin octamer binding element was 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotide (1ng) was incubated with HeLa nuclear extract (20μg), in the presence of 1μg of poly d(I-C).

Several complexes were formed between the probe and HeLa cell extract, in the absence of competitor DNA (lane 1). One of these complexes, indicated by an arrow, was shown to be specific as it was competed by increasing concentrations (10, 100 and 200ng) of homologous oligonucleotide (Specific) (lanes 2-4). Increasing concentrations (10, 100 and 200ng) of oligonucleotides containing either 5' AT-rich (5', lanes 5-7) or 3' T-Rich (3', lanes 8-10) sequences from the rPPT-A promoter also competed for complex formation. Increasing concentrations (100 and 200ng) of a non-specific oligonucleotide (lanes 11 and 12) failed to compete for complex formation.

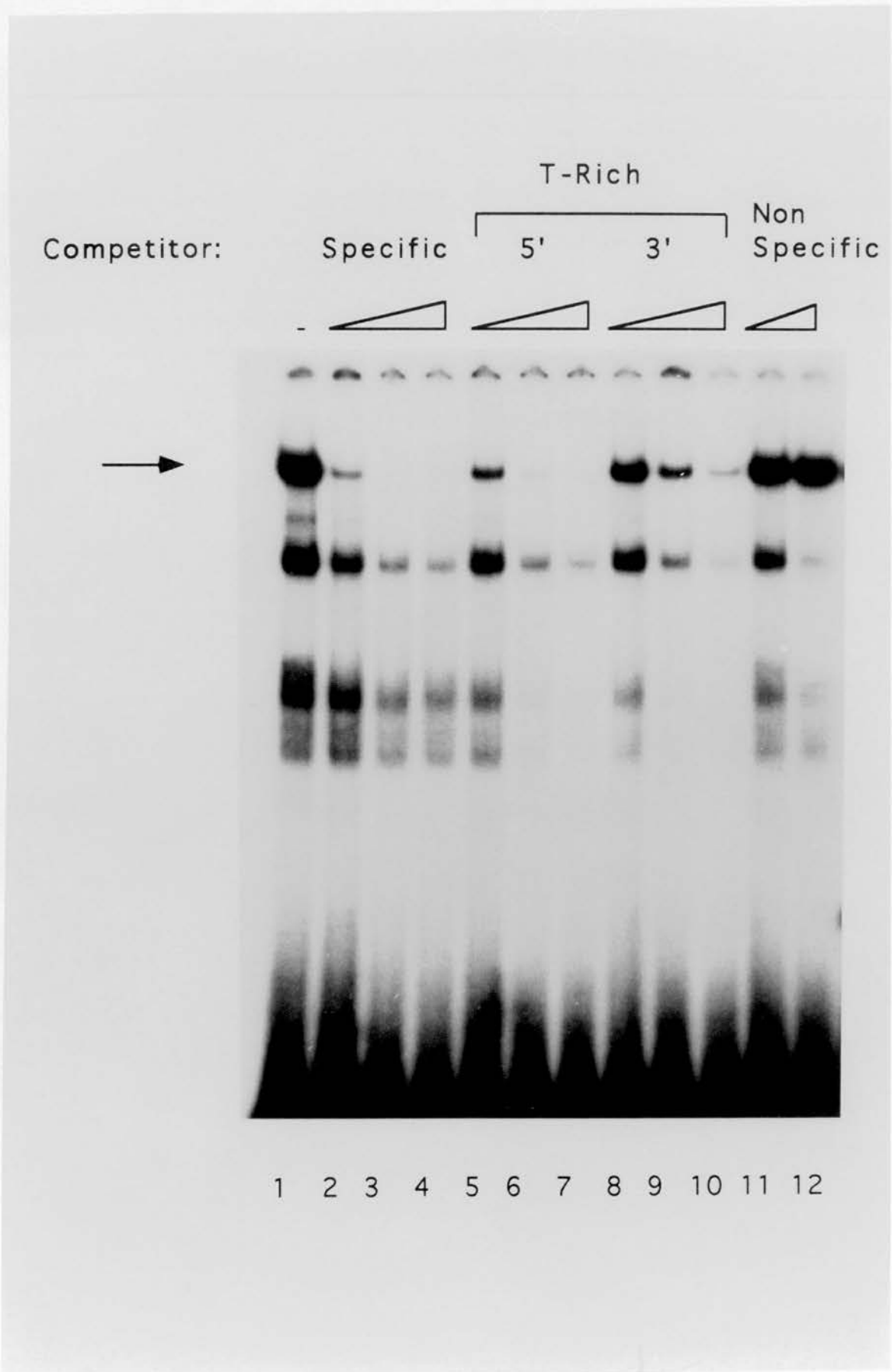
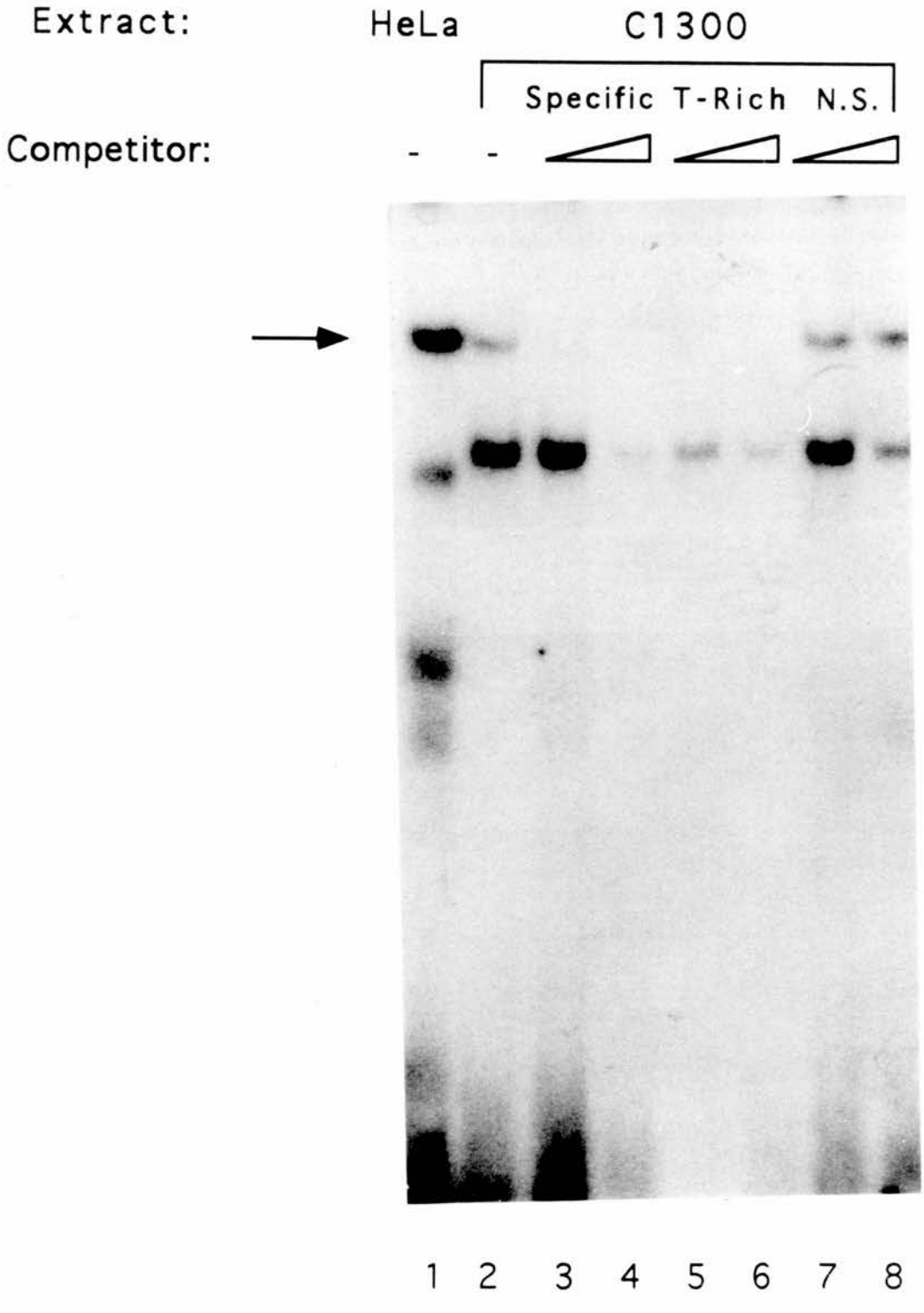

T-Rich
Competitor:
Specific
5'
3'
Non
Specific
-
1 2 3 4 5 6 7 8 9 10 11 12

Figure 18. Electrophoretic mobility shift assay using the immunoglobulin octamer binding element and extract from C1300 and HeLa cells.

A double stranded oligonucleotide containing the immunoglobulin octamer binding element was 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotide (1ng) was incubated with either HeLa nuclear extract (20μg) or C1300 whole cell extract (20μg), in the presence of 1μg of poly d(I-C).

Lane 1 contains HeLa cell extract and lanes 2-8 contain C1300 cell extract. Several complexes were formed between the probe and HeLa cell extract (lane 1) and between the probe and C1300 cell extract (lane 2), in the absence of competitor DNA. One of the complexes formed with C1300 cell extract, indicated by an arrow, was shown to be specific as it was competed by increasing concentrations (10 and 100ng) of homologous oligonucleotide (Specific) (lanes 3 and 4). Increasing concentrations (10 and 100ng) of an oligonucleotide containing a 5' AT-rich sequence from the rPPT-A promoter, (T-Rich) (lanes 5 and 6), also competed for complex formation and increasing concentrations (10 and 100ng) of a non-specific oligonucleotide, (N.S.) (lanes 7 and 8), failed to compete.


Extract:
HeLa
C1300
Specific
T-Rich
N.S.
Competitor:
-
-
1 2 3 4 5 6 7 8

similar position to that formed between C1300 cell extract and the same oligonucleotide (lane 2). The formation of the complex generated by C1300 cell extract was competed by increasing concentrations of both homologous, (Specific) (lanes 3 and 4), and 5' AT-rich, (T-Rich) (lanes 5 and 6), competitor oligonucleotides. Increasing concentrations of non-specific oligonucleotide, (N.S.) (lanes 7 and 8), failed to compete for complex formation.

Therefore, a region of the rPPT-A promoter containing a six out of eight base pair match to the octamer binding protein consensus sequence has affinity for octamer binding proteins present in both HeLa nuclear and C1300 whole cell extracts.

3.6.1.12. Element 11.

It has previously been shown (Quinn and McAllister, 1993) that one of the DNA strands from an element within the rPPT-A promoter, spanning nucleotides -761 to -741, binds a single-stranded sequence-specific DNA binding protein. By electrophoretic mobility shift assays it was demonstrated that this protein is present in extracts prepared from a series of adult rat tissues (brain, cerebellum, spinal cord and spleen) and in cultured adult rat DRG, but that it is absent from a series of neuronal and non-neuronal derived cell lines (Quinn and McAllister, 1993). A mechanism for binding of this protein to single-stranded DNA was proposed based on the fact that the region binding the single-stranded DNA binding protein can be modelled as the loop of a putative stem-loop structure. Further analysis was carried using an oligonucleotide, spanning nucleotides -790 to -738, which contains the entire sequence of the putative stem-loop structure. This longer oligonucleotide, in addition to binding single-stranded DNA binding proteins present in rat tissue extract, was also found to bind proteins from tissue culture cell lines when in the double-stranded form. Binding to the double-stranded form was not observed with rat tissue extract (McAllister *et al.*, 1993).

In order to investigate the binding of proteins from HeLa nuclear extract to this region of the promoter, DNase 1 footprinting analysis was carried out using the construct prPPT-βGal3 (Figure 6). Figure 19, lane 2 shows that HeLa nuclear extract generated a footprint (indicated as footprint 1) spanning nucleotides -779 to -755, as shown below;

$^{-779}$ CCCTTCTGCCTTCAGGGTGTGCCTG $^{-755}$

Interestingly, when a 30 fold excess of the oligonucleotide spanning nucleotides -790 to -738 was preincubated with the HeLa extract (lane 3), a different footprinting pattern was generated (indicated as footprint 2) which spans nucleotides -759 to -731, as shown below:

$^{-759}$ GCCTGGGAAGAAGCTGTAGGGGAACAAAA $^{-731}$

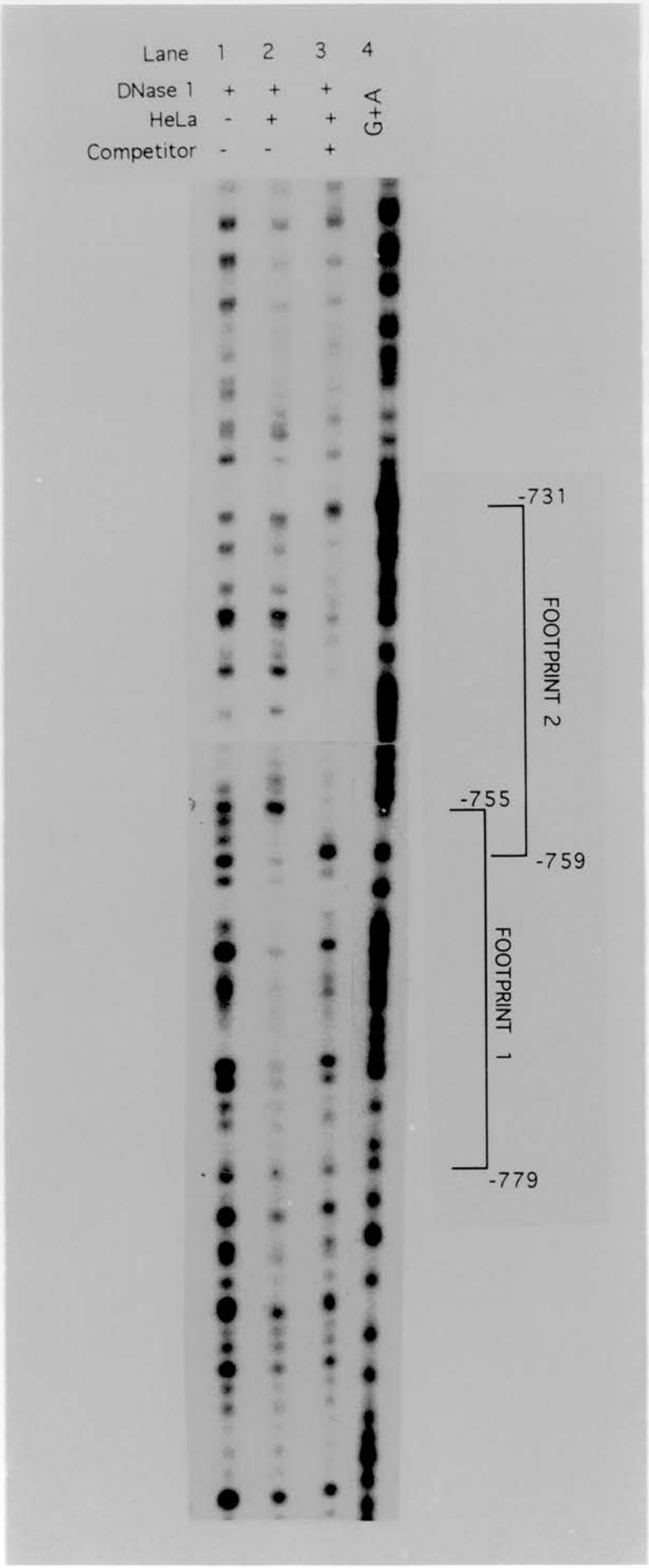
Figure 19. DNase 1 footprinting analysis of a region of the rPPT-A promoter which binds both single-stranded and double-stranded DNA binding proteins.

The construct prPPT-βGal3 was singly 5' end labelled with T4 polynucleotide kinase and 100ng of labelled probe was digested with DNase 1, in the absence or presence of HeLa nuclear extract.

Lane 1 shows the pattern of DNA digestion generated by 1.5 units of DNase 1. In addition to DNase 1 (30 units), HeLa nuclear extract (1mg) was added in lanes 2 and 3. In lane 3 an oligonucleotide (3μg), spanning nucleotides -790 to -738 of the rPPT-A promoter, was preincubated with the HeLa nuclear extract. Lane 4 shows the G+A Maxam Gilbert sequencing reaction of the DNA probe.

A footprinted region was generated by HeLa nuclear extract (lane 2) which, from the sequence in lane 4, spans nucleotides -779 to -755, indicated as footprint 1. The presence of the oligonucleotide in lane 3 resulted in the generation of a new footprint spanning nucleotides -759 to -731, indicated as footprint 2.

These two footprints cover a region of the rPPT-A promoter which has previously been shown to bind both single-stranded and double-stranded DNA binding proteins and has also been proposed to form a stem-loop structure (Quinn and McAllister, 1993; McAllister *et al.*, 1993).


Lane 1 2 3 4
DNase 1 + + +
HeLa - + +
Competitor - - +
G+A
-731
FOOTPRINT 2
-755
-759
FOOTPRINT 1
-779

6.1.13. Element 12.

A region of the rPPT-A promoter, located between the TATA box and the major transcriptional start site, termed the minimal promoter element, displays sequence homology with a similarly positioned element from the pituitary-specific pro-opiomelanocortin (POMC) gene promoter. The POMC element spans nucleotides -3 to -15 and specifically binds a factor termed PO-B which acts as a transcriptional activator, facilitating POMC gene expression (Riegel *et al.*, 1990). A comparison of the sequences around the POMC and rPPT-A elements is shown below;

PPT $^{-26}$ AA-tAagGcGAAGcAgGaG-CAGGGAC $^{-2}$

POMC $^{-23}$ AAgaAgaGaGAAG-A-GtGaCAGGGAC $^{+2}$

The two elements show 65% sequence homology suggesting that they may bind similar or related proteins.

In order to further analyse the rPPT-A minimal promoter element and to investigate any potential relationships with the POMC element an oligonucleotide (termed 168/169, Figure 10) was made, spanning nucleotides -20 to +4 of the rPPT-A promoter, for use in electrophoretic mobility shift assays. Figure 20, lane 1 shows that a complex (indicated by an arrow) was formed between the oligonucleotide 168/169 and HeLa nuclear extract. This complex was shown to be specific as it was competed by increasing concentrations of homologous DNA competitor, (Specific) (lanes 2-4), and not by increasing concentrations of a heterologous DNA competitor, (Octamer, lanes 5-7).

In order to investigate the possible relationship between PO-B and the protein(s) binding to the rPPT-A minimal promoter element, electrophoretic mobility shift assays were carried out using an oligonucleotide corresponding to nucleotides -23 to +5 of the POMC promoter, (a gift from A.Riegel, Department of Pharmacology, Georgetown University, Washington DC) (termed POB, Figure 10). Figure 21, lane 1 shows a complex (indicated by an arrow) formed between the POB probe and HeLa nuclear extract. This complex was specifically competed by increasing concentrations of homologous competitor, (Specific) (lanes 2-4), but not by increasing concentrations of the oligonucleotide 168/169 (lanes 5-7). This suggests that the proteins binding to the POMC and rPPT-A elements are not related. Further support for this observation comes from the finding that the complexes formed between oligonucleotides 168/169 and POB and various protein extracts have different mobilities. Figure 22 shows the complexes (indicated as arrow 3) formed between the POB probe and HeLa nuclear extract in lane 1, C1300 whole cell extract in lane 2 and AtT20 whole cell extract in lane 3. Lanes 4-6 show the complexes (indicated as arrow 2) formed between these same

Figure 20. Electrophoretic mobility shift assay using the rPPT-A minimal element and extract from HeLa cells.

A double stranded oligonucleotide (168/169) containing rPPT-A sequence, spanning nucleotides -20 to +4 relative to the major transcriptional start site, was 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotide (1ng) was incubated with HeLa nuclear extract (20µg), in the presence of 1µg of poly d(I-C).

Complexes were formed between the probe and HeLa cell extract (lane 1), in the absence of competitor DNA. One of these complexes, indicated as an arrow, was shown to be specific as it was competed by increasing concentrations (25, 50 and 100ng) of homologous oligonucleotide (Specific) (lanes 2-4). Increasing concentrations (25, 50 and 100ng) of an oligonucleotide containing the immunoglobulin octamer binding element, (Octamer) (lanes 5-7), failed to compete for complex formation.

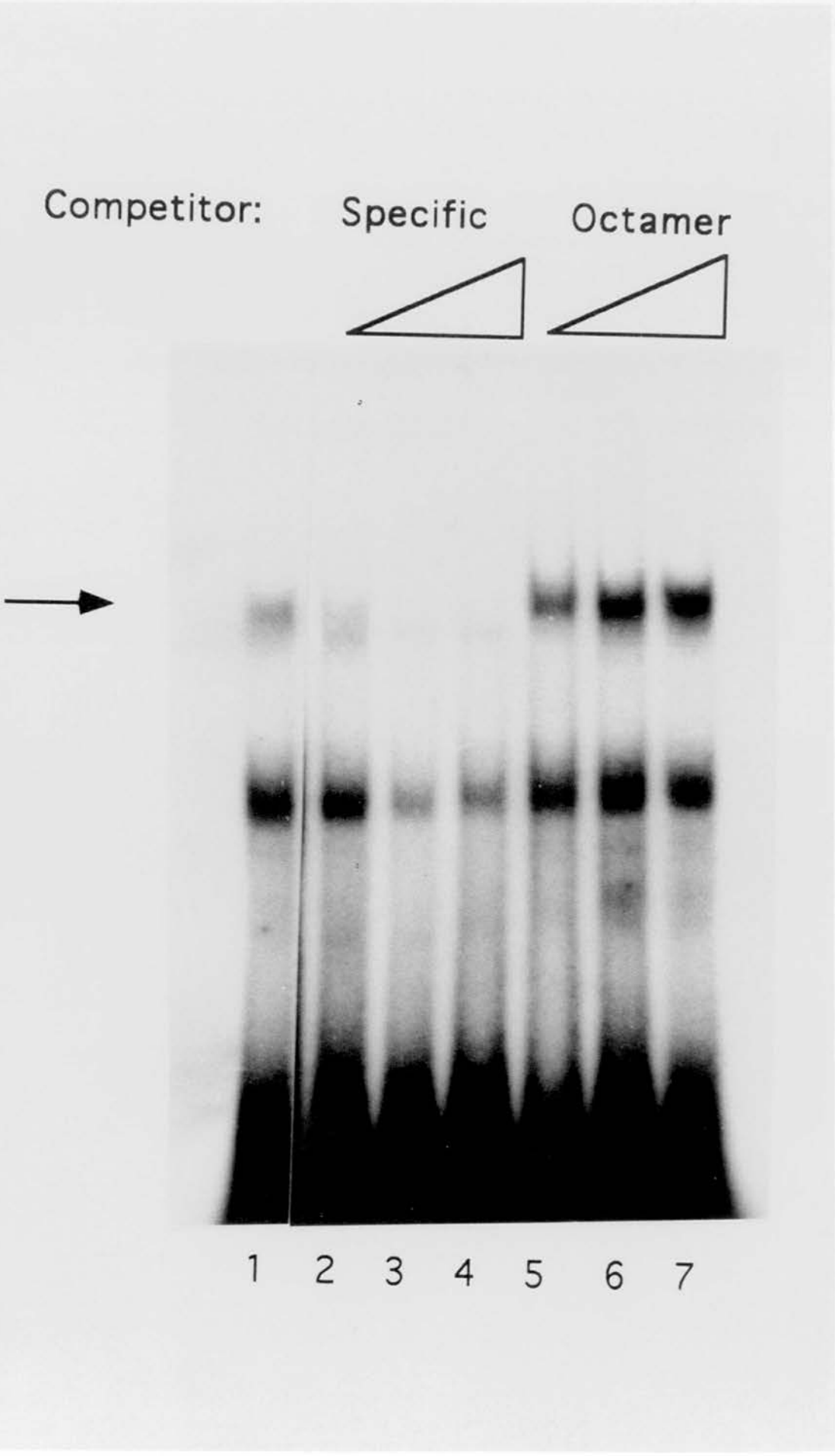

Competitor:
Specific
Octamer
1
2
3
4
5
6
7

Figure 21. Electrophoretic mobility shift assay using an element from the POMC promoter and extract from HeLa cells.

A double stranded oligonucleotide containing an element from the POMC promoter, termed POB, spanning nucleotides -23 to +5 relative to the major transcriptional start site was 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotide (1ng) was incubated with HeLa nuclear extract (20μg), in the presence of 1μg of poly d(I-C).

Complexes were formed between the probe and HeLa cell extract, in the absence of competitor DNA (lane 1). One of these complexes, indicated as an arrow, was shown to be specific as it was competed by increasing concentrations (25, 50 and 100ng) of homologous oligonucleotide (Specific) (lanes 2-4). Increasing concentrations (25, 50 and 100ng) of an oligonucleotide (168/169) containing sequence from the rPPT-A promoter (lanes 5-7) failed to compete for complex formation.

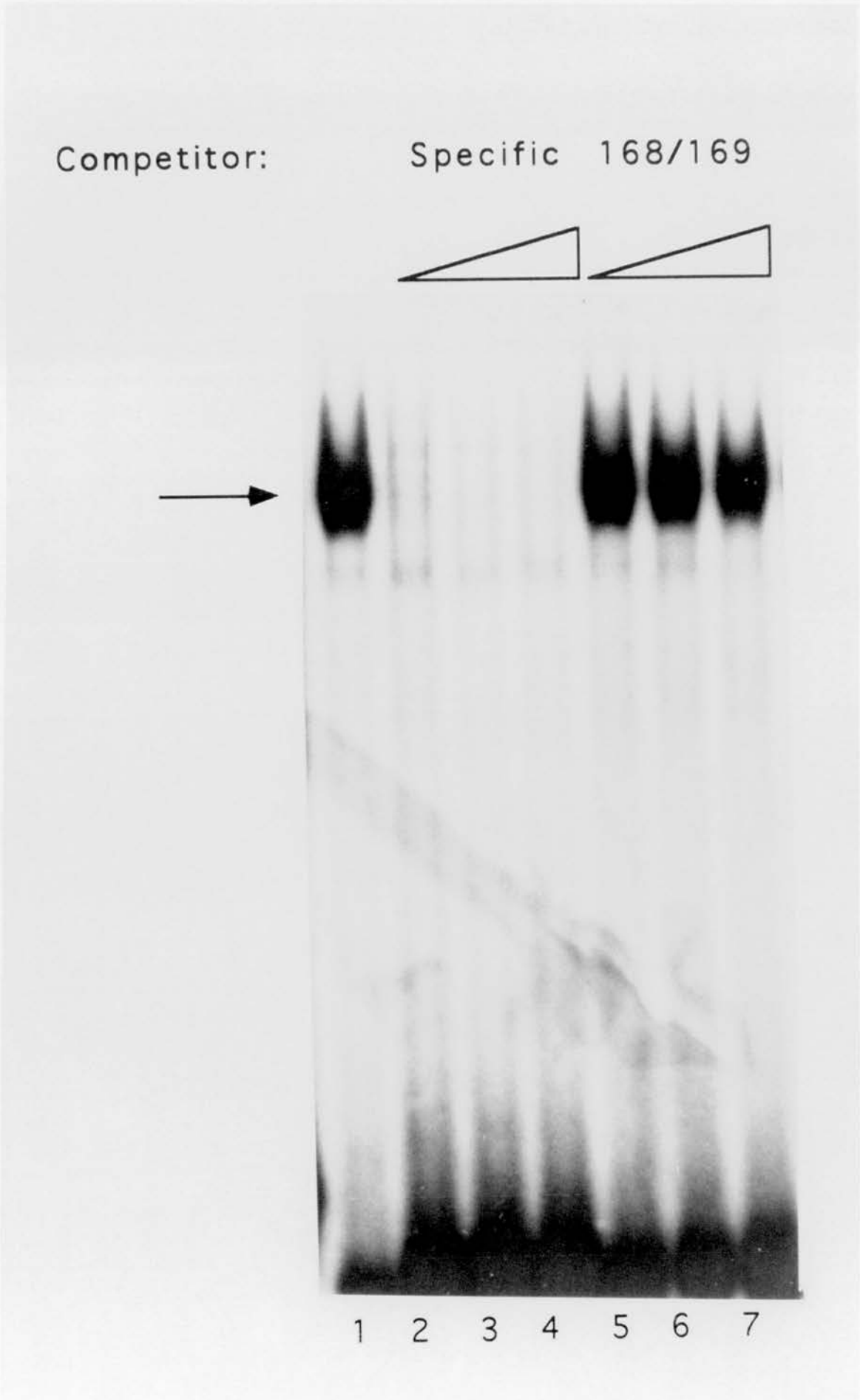
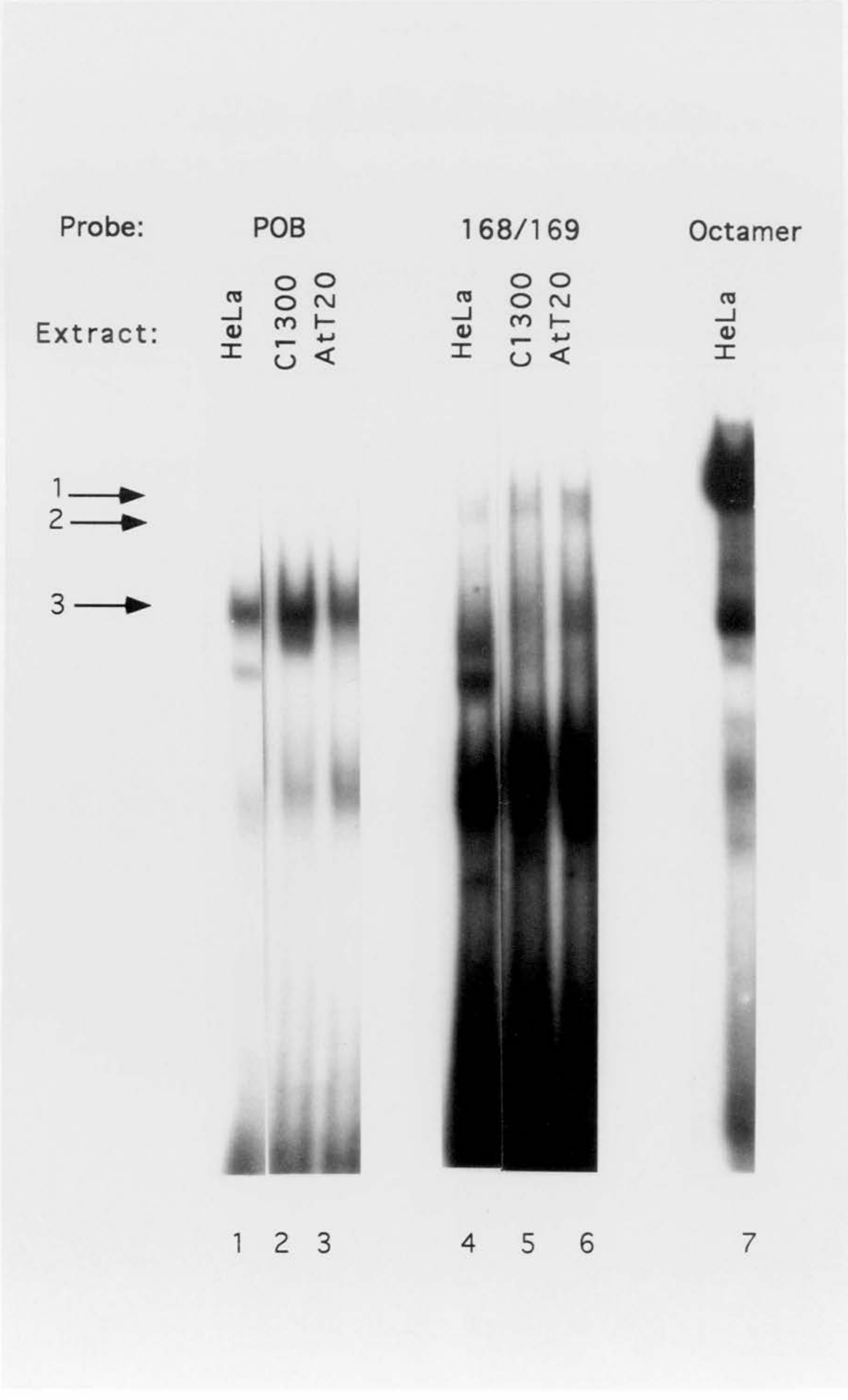

Competitor:
Specific 168/169
1 2 3 4 5 6 7

Figure 22 Electrophoretic mobility shift assay using either the rPPT-A minimal promoter element, an element from the POMC promoter or the immunoglobulin octamer binding element and extract from HeLa, C1300 and AtT20 cells.

Double stranded oligonucleotides containing either rPPT-A sequence spanning nucleotides -20 to +4 (168/169), an element from the POMC promoter spanning nucleotides -23 to +5 (POB) or the immunoglobulin octamer binding element (Oct) were 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotides (1ng) were incubated with either HeLa nuclear extract, C1300 or AtT20 whole cell extracts (20μg each), in the presence of 1μg of poly d(I-C).

Complexes were formed between the POB probe and HeLa nuclear extract, C1300 whole cell extract and AtT20 whole cell extract (lanes 1-3 respectively). Arrow 3 indicates the position of the specific complexes, as determined from previous electrophoretic mobility shift assays (Figure 21). Complexes were also formed between the 168/169 probe and HeLa nuclear extract, C1300 whole cell extract and AtT20 whole cell extract (lanes 4-6 respectively). Arrow 2 indicates the position of specific complexes, as determined from previous electrophoretic mobility shift assays (Figure 20). In Lane 7 the specific complex (as determined in previous electrophoretic mobility shift assays Figure 16, lane 14) formed between HeLa nuclear extract and the Oct probe is indicated by arrow 1.


Probe:
POB
168/169
Octamer
Extract:
HeLa
C1300
AtT20
HeLa
C1300
AtT20
HeLa
1
2
3
1 2 3 4 5 6 7

extracts and the rPPT-A 168/169 probe. The complex (indicated as arrow 1) generated by the ubiquitous Oct-1 protein present in HeLa nuclear extract and an octamer containing oligonucleotide (Oct, Figure 10) is indicated in lane 7. From this figure it is clear that the complexes formed between the probes from the POMC promoter and the rPPT-A promoter have very different mobilities, suggesting that distinct proteins bind to these two promoter elements.

Interestingly, DNase 1 footprinting analysis carried out using the construct pSM1/Q5 (Figure 6) and HeLa nuclear extract also revealed protein binding activity within this region of the rPPT-A promoter (i.e. between the TATA box and the transcriptional start site). Figure 23 (lanes 2 and 3) shows that in addition to a strong footprint covering nucleotides -67 to -47 (E-box 1, element 1), HeLa nuclear extract generated a footprint (indicated by arrows) covering nucleotides -16 and -17, shown below in bold type:

-20 GCG**AA**GC -14

In addition to the footprint, HeLa nuclear extract also generated a hypersensitive site covering a dC nucleotide at position -19. This nucleotide was not cleaved by DNase 1 in the absence of protein, but was cleaved in the presence of protein (compare lane 1 with lanes 2 and 3).

Footprints generated over this region of the rPPT-A promoter were only occasionally observed. This finding was reflected in electrophoretic mobility shift assays. Complexes formed between the 168/169 oligonucleotide and protein extracts were not routinely observed, in some cases the specific shift was very weak or not present at all. This suggests that the protein/DNA interaction(s) in this region of the promoter are in some way unstable, perhaps due to the lability of the protein(s) responsible for complex formation.

3.6.1.14. Elements 13 to 16.

By use of the exonuclease protection assay multiple and specific DNA-protein interaction sites have previously been identified within a 3.8 kb DNA fragment of the rPPT-A promoter (Quinn, 1992). One of the sites of interaction was mapped to a region 3' of the transcriptional start site, approximately at the intron 1/exon 1 boundary. The proteins binding to this site displayed neuronal specificity in their tissue distribution. Proteins from rat cerebellar, hippocampal, spinal cord and PC12 whole cell extracts displayed binding activity, whereas binding was not observed by proteins present in rat spleen extract or in the non-neuronal derived cell lines HeLa and 293 (Quinn, 1992).

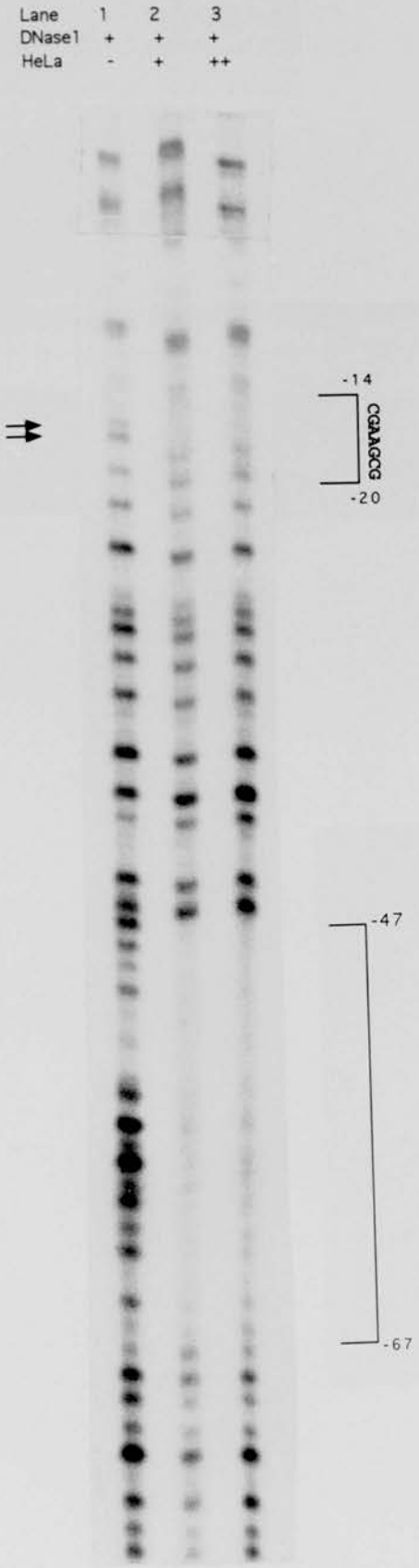
In order to further investigate this apparently neuronal specific protein/DNA interaction site, DNase 1 footprinting assays were carried out using a DNA fragment

Figure 23. DNase 1 footprinting analysis of a region of the rPPT-A promoter located between the TATA box and transcriptional start site.

The construct pSM1/Q5 was singly 5' end labelled with T4 polynucleotide kinase and 100ng of labelled probe was digested with DNase 1, in the absence or presence of HeLa nuclear extract.

Lane 1 shows the pattern of digestion generated by 1.5 units of DNase 1. In addition to DNase 1 (30 units), increasing amounts of HeLa nuclear extract (0.5 and 1mg) were added in lanes 2 and 3.

Two footprinted regions were generated by HeLa nuclear extract (lanes 2 and 3). The lower footprint corresponds to the E-box 1 element spanning nucleotides -67 to -47. The upper footprint, indicated by arrows, spans nucleotides -16 and -17 (dAs) and nucleotide -19 (dC) appears to represent a hypersensitive site.


Lane 1 2 3
DNase1 + + +
HeLa - + ++
-14
CGAAGCG
-20
-47
-67

from the rPPT-A promoter, derived from the construct pSM1'/Q5, with HeLa and rat cerebellar nuclear extracts.

A suitable substrate for DNase 1 footprinting assays was generated by digestion of the construct pSM1'/Q5 (Figure 6) in the polylinker with the restriction enzymes *Eco*RI and *Hind*III. This generated a DNA fragment of approximately 650 bp, containing sequence spanning nucleotides -160 to +447 from the rPPT-A promoter. This DNA fragment was purified from a low melting point agarose gel and then digested with the restriction enzyme *Dde* I. A resulting DNA fragment of approximately 500 bp, termed pSM1'/Q5/*Dde* I (Figure 6), which contains nucleotides -35 to +447 from the rPPT-A promoter, was purified and subsequently used in DNase 1 footprinting assays.

Figure 24 shows the result of a DNase 1 footprinting assay using pSM1'/Q5/*Dde* I with HeLa and rat cerebellar nuclear extracts. In lane 1 the pattern of DNase 1 digestion in the absence of protein extract is shown. In addition to DNase 1, HeLa and rat cerebellar (indicated as Cer.) extracts were added in lanes 2 and 3 respectively. It is clear that the two extracts generated different patterns of footprints. The cerebellar generated footprints have been termed elements 13 and 16 and the HeLa generated footprints have been termed elements 14 and 15.

Element 13.

Cerebellar nuclear extract (Figure 24, lane 3) generated a footprint spanning nucleotides +55 to +89, as shown below:

+55 CGCCCAG**CAAGTG**CG**CACCTG**CGGAGCATCACCGG +89

(Nucleotide +89 is a hypersensitive site.)

This element contains two E-box motifs (consensus sequence CANNTG), as shown in bold type. The **CAAGTG** motif is termed E-box 4 and the **CACCTG** motif is termed E-box 5.

Element 16.

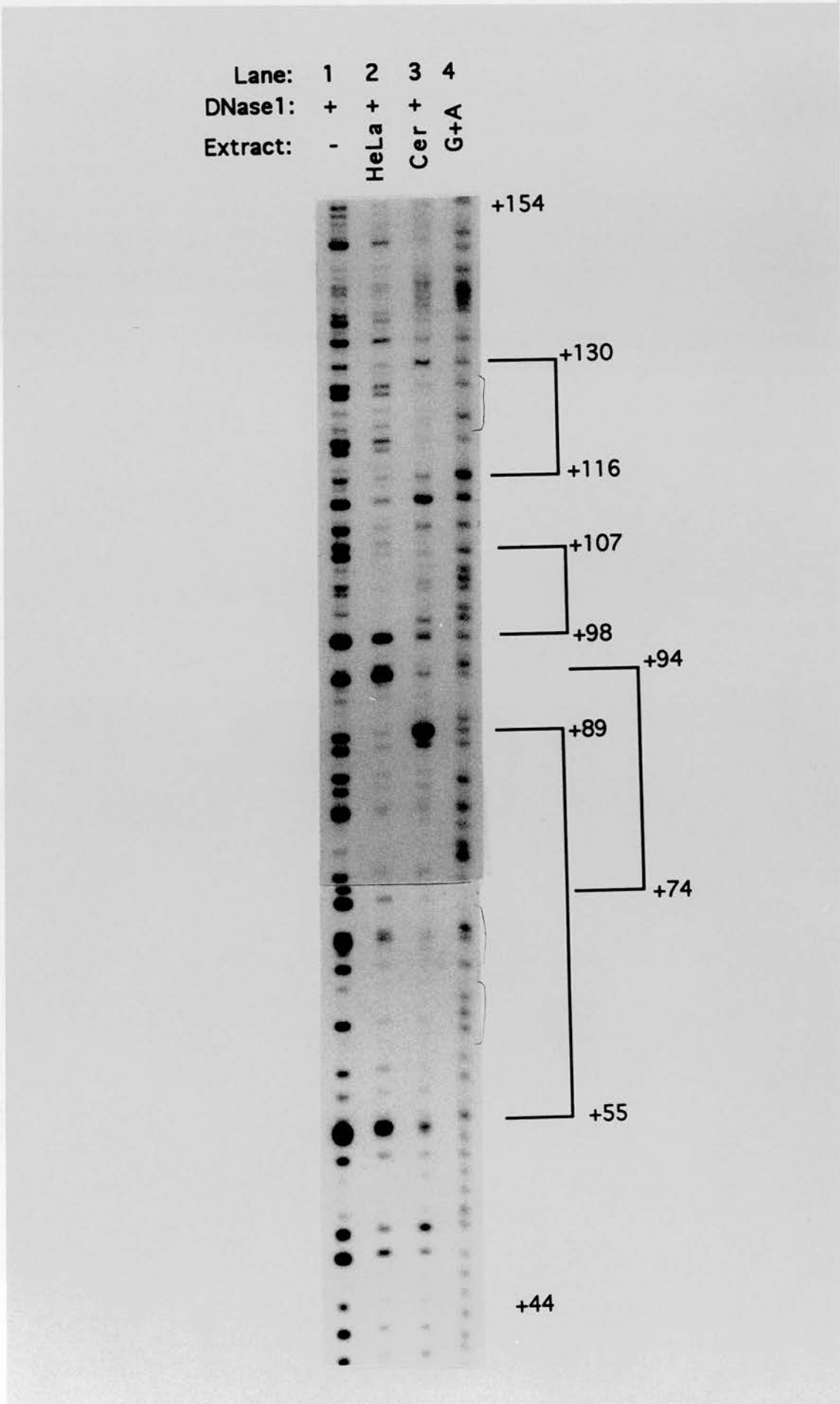
Cerebellar extract (Figure 24, lane 3) also generated a footprint spanning nucleotides +116 to +130, as shown below:

+116 GCCTGC**CATCTG**CCG +130

This elements also contains an E-box motif (shown in bold type), termed E-box 6.

Figure 24. DNase 1 footprinting analysis of a region of the rPPT-A promoter which shows tissue specificity in its protein binding activity.

The construct pSM1/Q5 was digested with the restriction enzymes *Eco*R I, *Hind* III and *Dde* I to produce a DNA fragment of approximately 500 bp in length. This fragment was singly 5' end labelled with T4 polynucleotide kinase and 100ng was digested with DNase 1, in the absence or presence of either HeLa or rat cerebellar nuclear extracts.
Lane 1 shows the pattern of digestion generated by 2.5 units of DNase 1. In addition to DNase 1 (30 units in lane 2 and 15 units in lane 3), HeLa (1mg) and rat cerebellar extract (Cer.) (1mg) were added in lanes 2 and 3 respectively . Lane 4 shows the G+A Maxam Gilbert sequencing reaction of the DNA probe.
The neuronal and non-neuronal derived extracts generated different patterns of protection over this region of the promoter.
Two footprints were generated by HeLa nuclear extract (lane 2) which, from the sequence in lane 4, span nucleotides +74 to +94 and +98 to +107. The footprint spanning nucleotides +74 to +94 is contained within exon 1 and the footprint spanning nucleotides +98 to +107 is located over the intron 1/exon 1 boundary. When this footprint is extended to nucleotide +108 it displays a five out of seven base pair match to an AP-1 consensus sequence and so is termed HeLa AP-1.
Two footprints were also generated by rat cerebellar extract (lane 3) which, from the sequence in lane 4, span nucleotides +55 to +89 and +116 to +130, as indicated. The footprint spanning nucleotides +55 to +89 is located in exon 1 and the footprint spanning nucleotides +116 to +130 is located in intron 1. Two E-box motifs, termed E-box 4 and 5 are contained within the footprint spanning nucleotides +55 to +89 and one E-box motif, termed E-box 6, is contained within the footprint spanning nucleotides +116 to +130.


Lane: 1 2 3 4
DNase1: + + + +
Extract: - HeLa Cer G+A
+154
+130
+116
+107
+98
+94
+89
+74
+55
+44

Element 14.

When HeLa extract (lane 2) was added a footprint was generated covering nucleotides +74 to +94, as shown below:

$^{+74}$ TGCGGAGCATCACCGGGTCCG $^{+94}$

Element 15.

HeLa nuclear extract also generated a footprint spanning nucleotides +98 to +107, as shown below:

$^{+98}$ GCAGTGAGTA $^{+107}$

When the sequence of this footprint is extended to nucleotide +108 (C), a region of sequence, termed HeLa AP-1, is revealed which displays a five out of seven base pair match to an AP-1 consensus sequence (TGA G/C TCA).

Therefore, the neuronal and non-neuronal derived protein extracts protected distinct yet overlapping sequences in a region of the rPPT-A promoter spanning nucleotides +55 to +130. These results provide support for the finding by Quinn (1992) of tissue specific binding in this region and additionally allow the sites of protein/DNA interaction to be more accurately mapped.

The protected elements are located both in exon 1, intron 1 and over the intron 1/exon 1 boundary. Two of the protected regions, the neuronal generated footprint from +55 to +89 and the non-neuronal derived footprint from +74 to +94, overlap and are contained in exon 1. Intron 1 starts at +101 therefore placing the second HeLa cell-generated footprint, from +98 to +107, over the intron 1/exon 1 boundary. Finally the second footprint generated by cerebellar extract covering nucleotides +116 to +130 is placed in intron 1.

3.6.1.15. Element 17.

DNase 1 footprinting was carried out using the construct pSM1'/Q5 (Figure 6) with HeLa, rat cerebellum and total rat brain nuclear extracts (Figure 25). A footprint, located 3' of the major transcriptional start site, was generated spanning nucleotides +268 to +290, as shown below;

$^{+268}$ AGTGGGTAGGGGGCTGGGACGTT $^{+290}$

This element is rich in dG nucleotides and so is termed a 3' G-rich element. It can be seen from Figure 25 that HeLa nuclear extract (lane 2) generated a very strong footprint covering this G-rich element and that rat cerebellar extract, (indicated as Cer.) (lane 3), generated a slightly weaker footprint while rat total brain extract, (indicated as Brn.) (lane 4), did not appear to generate any footprint over this region.

Figure 25. DNase 1 footprinting analysis of a region of the rPPT-A promoter encompassing U.D. 2 and a G-rich sequence.

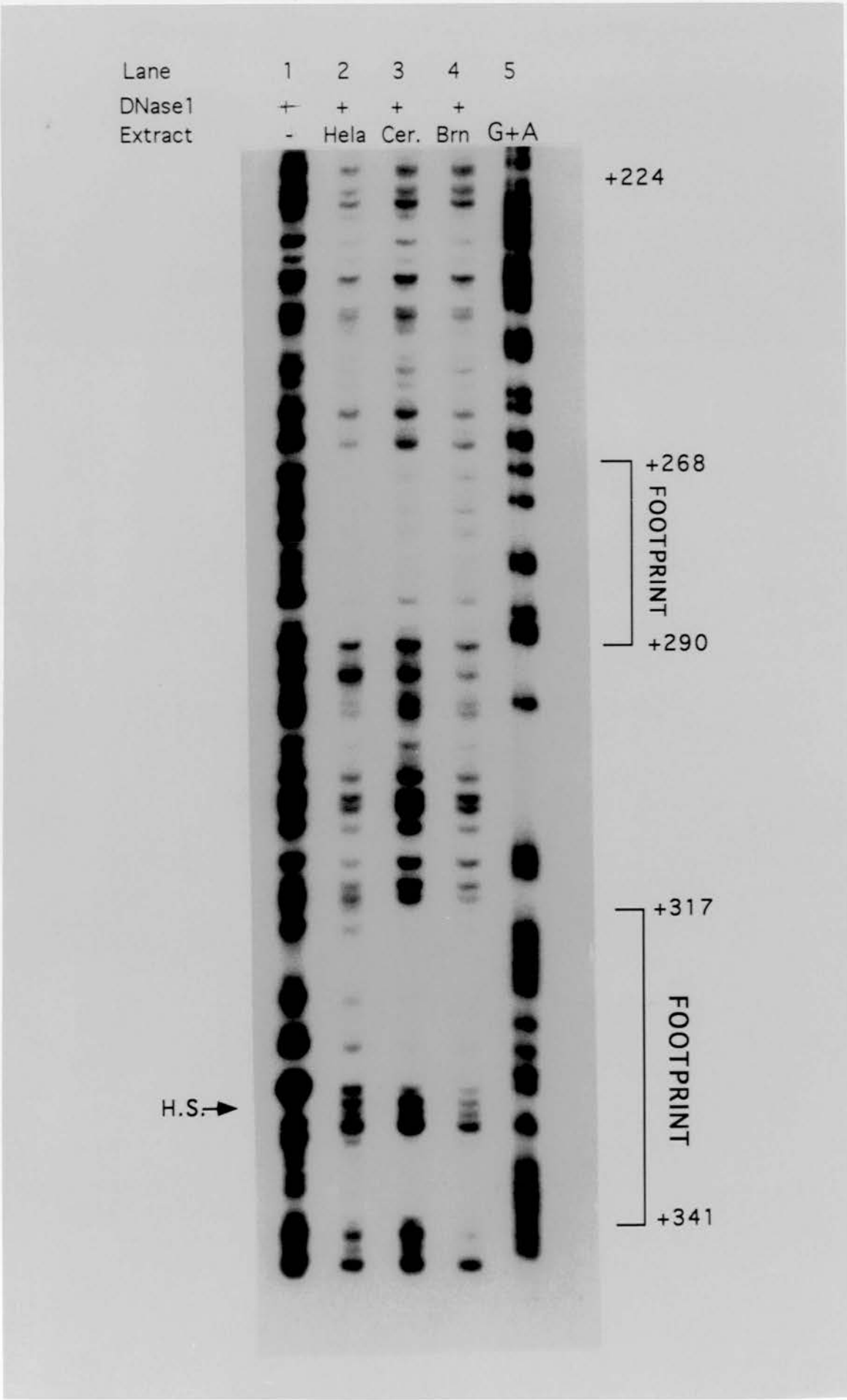
The construct pSM1'/Q5 was singly 5' end labelled with T4 polynucleotide kinase and 100ng of labelled probe was digested with DNase 1, in the absence or presence of either HeLa, rat cerebellar or total rat brain nuclear extracts.

Lane 1 shows the pattern of digestion generated by 1.5 units of DNase 1. In addition to DNase 1 (30 units), HeLa nuclear extract (1mg) was added in lane 2, rat cerebellar extract, (Cer.) (1mg), in lane 3 and rat brain extract, (Brn) (1mg), in lane 4. Lane 5 shows the G+A Maxam Gilbert sequencing reaction of the DNA probe.

Two footprints were generated by the extracts (lanes 2 to 4) which, from the sequence in lane 5, span nucleotides +317 to +341 and +268 to +290.

The footprint generated by HeLa nuclear extract (lane 2), spanning nucleotides +317 to +341, was much weaker than those generated by cerebellar (lane 3) or brain (lane 4) extracts. All three extracts additionally generated hypersensitive sites, (indicated as HS), covering nucleotides +331 and +334. This footprinted region is termed U.D. 2.

A footprint was also generated by HeLa nuclear extract (lane 2) and rat cerebellar extract (lane 3), as indicated, spanning nucleotides +268 to +290. The total rat brain extract (lane 4) did not appear to generate any footprint over this region. This footprinted region covers a sequence rich in dG nucleotides and therefore is termed 3' G-rich.


Lane
1
2
3
4
5
DNase1
+
+
+
+
Extract
-
Hela
Cer.
Brn
G+A
+224
+268
FOOTPRINT
+290
+317
FOOTPRINT
+341
H.S.

In order to further analyse the binding of proteins to this 3' G-rich element an oligonucleotide, spanning nucleotides +264 to +291, containing the footprinted region and flanking sequence (3' G-rich, Figure 10) was made for use in electrophoretic mobility shift assays. Figure 26, lane 1 shows that a complex, indicated as an arrow, was formed between the 3' G-rich oligonucleotide and HeLa nuclear extract. This complex was shown to be specific as it was competed by increasing concentrations of homologous competitor, (Specific) (Figure 26, lanes 2-4 and Figure 27, lanes 1-3), but not by the heterologous competitor oligonucleotides, PPT AP-1 (Figure 10) in Figure 27, lanes 4-6 and 3' T-Rich (Figure 10) in Figure 26, lanes 5-7. For comparison, a specific complex formed between HeLa nuclear extract and an oligonucleotide, (PPT 'CRE', Figure 10), containing the 'CRE' element (element 3), spanning nucleotides -200 to -186, is shown in Figure 27, lane 7.

Attempts to carry out electrophoretic mobility shift assays using rat cerebellar and total brain extracts were not successful. This is consistent with the DNase 1 footprinting analysis in which weaker binding was displayed by these extracts (Figure 25, lanes 3 and 4).

3.6.1.16. Element 18.

DNase 1 footprinting analysis was carried out using the construct pSM1'/Q5 (Figure 6) with HeLa, total rat brain and rat cerebellar nuclear extracts. All three extracts generated footprints spanning nucleotides +317 to +341, (Figure 25, lanes 2-4), as shown below:

$^{+317}$ GCTCTTTGGCACGTCA**G**TAGCCTTC $^{+341}$

The pattern of footprinting generated by the different extracts show some differences. HeLa nuclear extract (lane 2) appeared to generate a weaker footprint than rat cerebellar, (indicated as Cer.) (lane 3), or brain extract, (indicated as Brn) (lane 4).

It can be seen from Figure 25 that this footprint actually appears to consist of two halves. One half covers nucleotides +317 to +332 and the second half covers nucleotides +334 to +341. In addition to the footprints, all three extracts generated a hypersensitive site (indicated as H.S.) covering a dG nucleotide at position +333 (shown in bold type above). This nucleotides was not cleaved by DNase 1 in the absence of protein (lane 1) but was cleaved in the presence of protein extract (lanes 2-4).

Inspection of the sequence reveals that the first half of the footprint contains a region which displays a six out of eight base pair match to a CRE/ATF binding motif (TGACGTCA), as shown in bold type below:

$^{+317}$ GCTCTTTG**GCACGTCA**GTAGCCTTC $^{+341}$

Figure 26. Electrophoretic mobility shift assay using the rPPT-A 3' G-rich element and extract from HeLa cells.

A double stranded oligonucleotide containing the rPPT-A 3' G-rich element, spanning nucleotides +264 to +291 relative to the major transcriptional start site, was 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotide (1ng) was incubated with HeLa nuclear extract (20µg), in the presence of 2.5µg of poly d(I-C). Several complexes were formed between the probe and HeLa nuclear extract (lane 1). One of these complexes, indicated as an arrow, was shown to be specific as it was competed by increasing concentrations (25, 50 and 100ng) of homologous oligonucleotide (Specific) (lanes 2-4). Increasing concentrations (25, 50 and 100ng) of an oligonucleotide containing a 3' T-rich sequence from the rPPT-A promoter, (T-Rich) (lanes 5-7), failed to compete for complex formation.

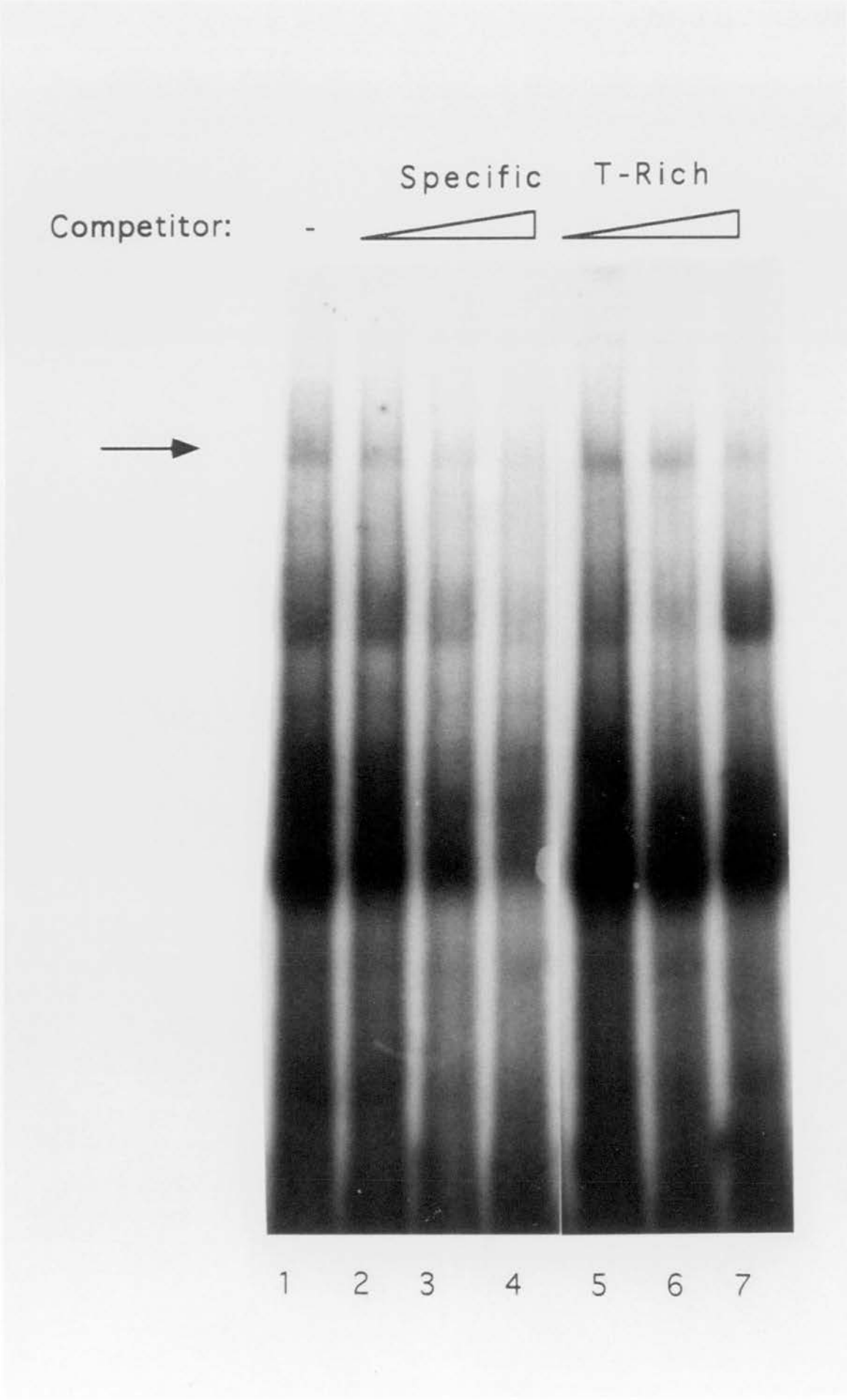
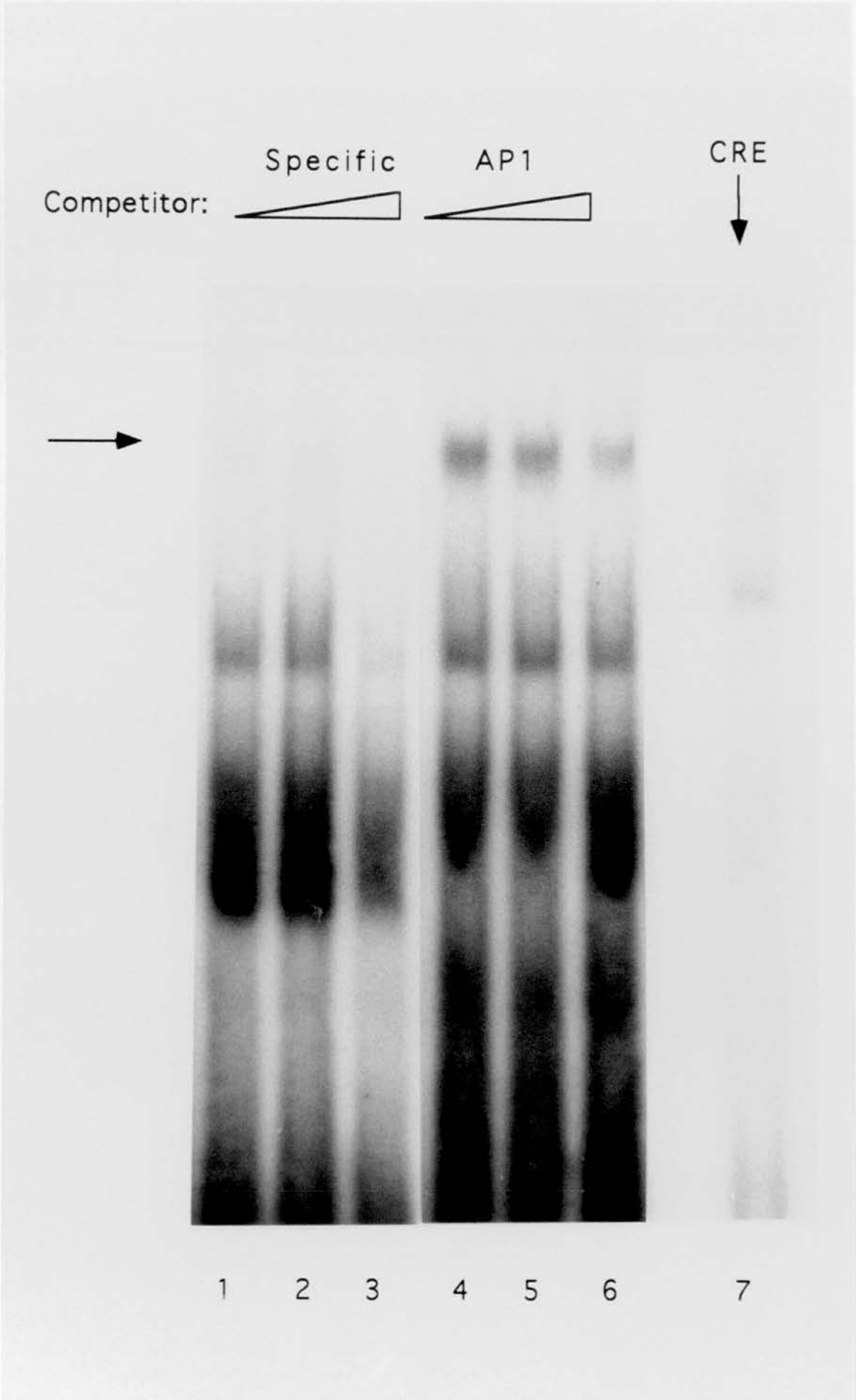

Specific
T-Rich
Competitor:
-
1
2
3
4
5
6
7

Figure 27. Electrophoretic mobility shift assay using the rPPT-A 3' G-rich element and extract from HeLa cells.

Double stranded oligonucleotides containing either the rPPT-A 3' G-rich element, spanning nucleotides +264 to +291, or the PPT 'CRE' element, spanning nucleotides -200 to -186, relative to the major transcriptional start site were 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotides (1ng) were incubated with HeLa nuclear extract (20µg), in the presence of 1µg of poly d(I-C) for 'CRE' and 2.5µg of poly d(I-C) for 3' G-rich.

Several complexes were formed between the 3' G-rich oligonucleotide and HeLa nuclear extract. One of the complexes was shown to be specific as it was competed by increasing concentrations (25, 50 and 100ng) of homologous oligonucleotide (Specific) (lanes 1-3). Increasing concentrations (25, 50 and 100ng) of an oligonucleotide containing an AP-1 element from the rPPT-A promoter, (AP1) (lanes 4-6), failed to compete for complex formation. Lane 7 shows the complex formed between the 'CRE' oligonucleotide and HeLa nuclear extract.


Specific
AP1
CRE
Competitor:
1
2
3
4
5
6
7

Electrophoretic mobility shift assays will be required in order to determine if this element does indeed bind a member of the CREB/ATF family of proteins.

As the identity of the proteins binding to the element have yet to be determined it has been termed unidentified element 2 (U.D. 2).

3.6.1.17. Element 19.

DNase 1 footprinting analysis was carried out using the construct pSM1'/Q5 (Figure 6) with HeLa, total rat brain and rat cerebellar nuclear extracts. Figure 28 shows that the pattern of footprinting over this region of the rPPT-A promoter appears to vary with different extracts.

Footprints were generated by the HeLa (lane 2) and total rat brain extracts (indicated as Brn., lane 4) but not by rat cerebellar extract (indicated as Cer., lane 3).

The footprint generated by HeLa nuclear extract spans nucleotides +372 to +390 and the brain generated footprint spans nucleotides +372 to +405, as shown below:

```
    +372 TGGGTTTTTGTTTTTGTTTTAATCTTGGTTTAGC +405
HeLa ————————————————————|
Brain ——————————————————————————————————|
```

Additionally, HeLa extract generated a hypersensitive site (indicated as H.S.) at nucleotide +381, (G) (shown in bold type), which was not generated by either of the brain derived extracts.

The sequence element footprinted by HeLa and total rat brain extracts is rich in dT nucleotides and so is termed a 3' T-rich element. In order to further characterise the binding of proteins from HeLa nuclear extract to this 3' T-rich element, electrophoretic mobility shift assays were carried out using an oligonucleotide containing the HeLa footprinted sequence and flanking regions, (3' T-rich, Figure 10). Figure 29, lane 1 shows that a complex (indicated by an arrow) was formed between the 3' T-rich oligonucleotide and HeLa nuclear extract. This complex was shown to be specific as it was competed by increasing concentrations of homologous DNA competitor, (Specific) (lanes 2 and 3), and not by increasing concentrations of a heterologous DNA competitor, (3' G-rich) (lanes 4-6).

When the sequence of the HeLa generated 3' T-rich footprint is compared to the 5' AT-rich sequence (element 8), in the reverse orientation, a significant degree of sequence homology (79%) is revealed:

Figure 28. DNase 1 footprinting analysis of a region of the rPPT-A promoter encompassing a T-rich sequence.

The construct pSM1'/Q5 was singly 5' end labelled with T4 polynucleotide kinase and 100ng of labelled probe was digested with DNase 1, in the absence or presence of either HeLa, rat cerebellar or total rat brain nuclear extracts.

Lane 1 shows the pattern of DNA digestion generated by 1.5 units of DNase 1. In addition to DNase 1 (30 units), HeLa nuclear extract (1mg) was added in lane 2, rat cerebellar extract (Cer.) in lane 3 and total rat brain nuclear extract (Brn) (1mg) in lane 4. Lane 5 shows the G+A Maxam Gilbert sequencing reaction of the DNA probe.

Footprints were generated by the HeLa extract (lane 2) and total rat brain extract (lane 4) which, from the sequence reaction (lane 5), span nucleotides +372 to +390 and +372 to +405 respectively. Rat cerebellar extract (lane 3) failed to generate any footprints over this region of the promoter. The sequence of the footprints are rich in dT nucleotides and are therefore termed 3' T-rich.

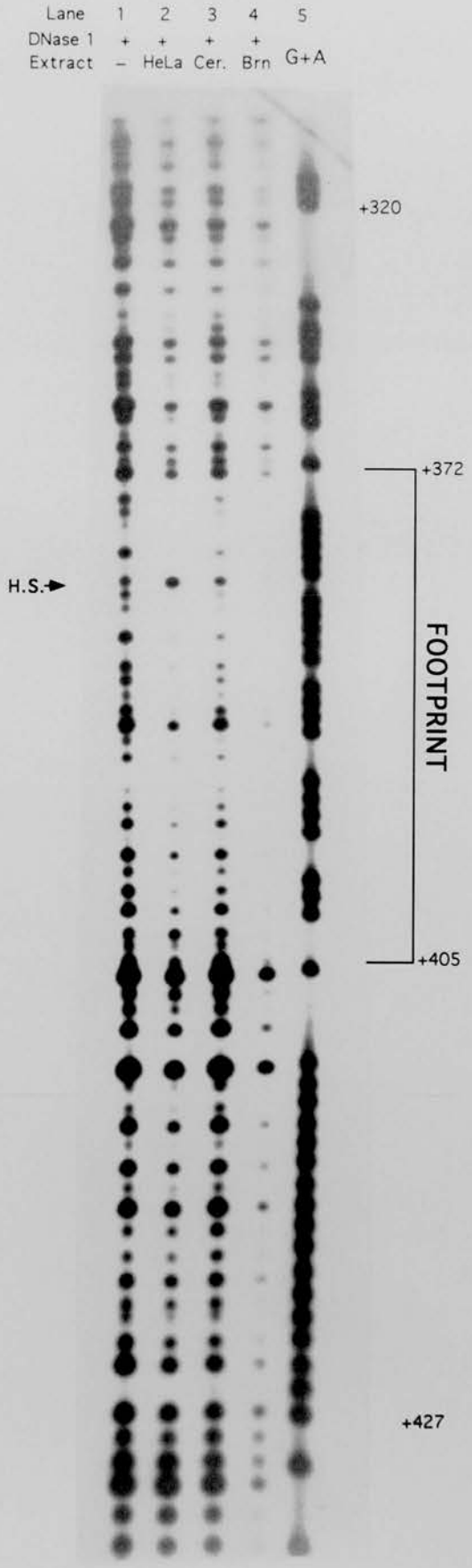
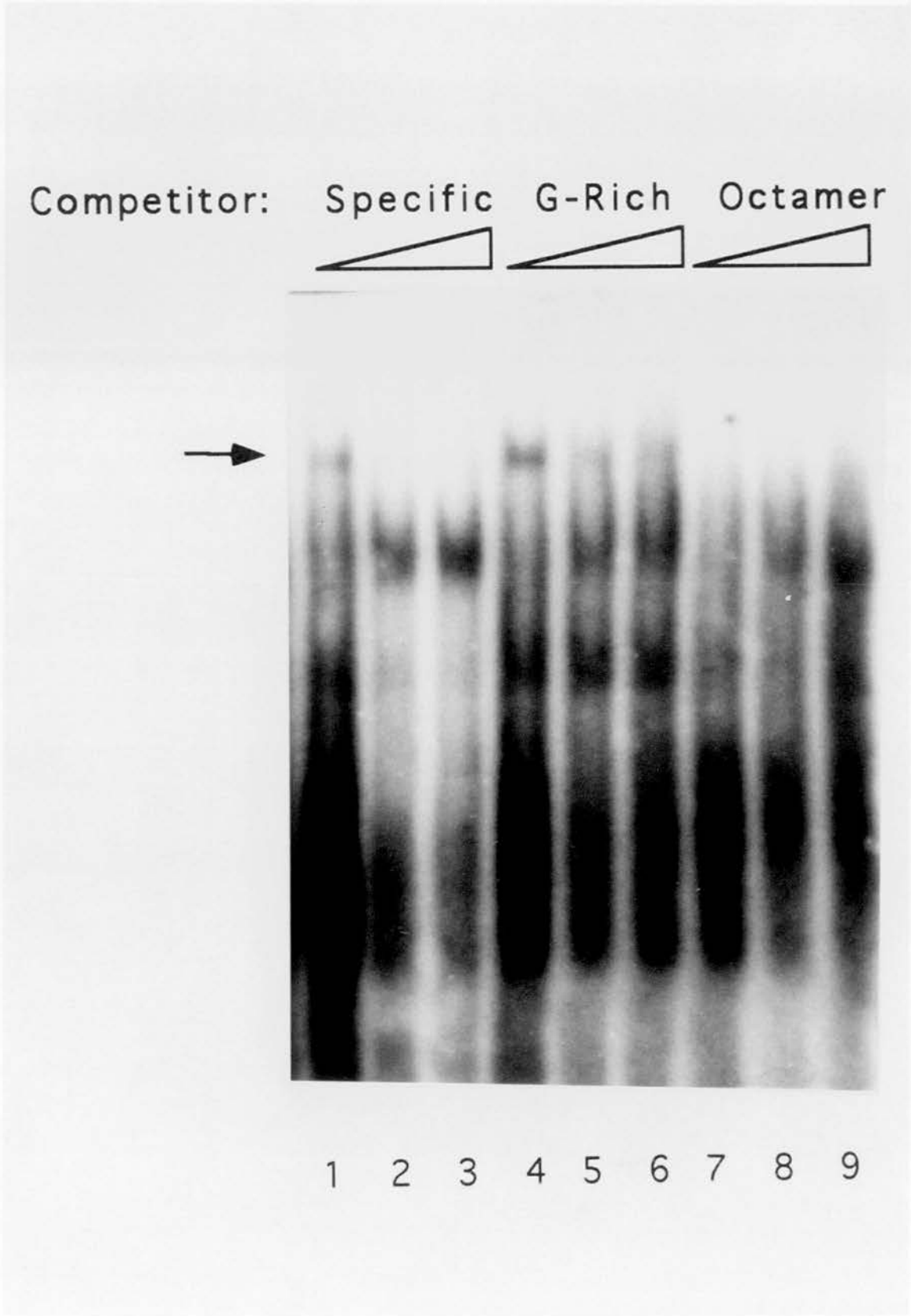

Lane 1 2 3 4 5
DNase 1 + + + +
Extract – HeLa Cer. Brn G+A
+320
+372
H.S.
FOOTPRINT
+405
+427

Figure 29. Electrophoretic mobility shift assay using the rPPT-A 3' T-rich element and extract from HeLa cells.

A double stranded oligonucleotide containing the rPPT-A 3' T-rich element, spanning nucleotides +369 to +395 relative to the major transcriptional start site, was 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotide (1ng) was incubated with HeLa nuclear extract (20µg), in the presence of 2.5µg of poly d(I-C). Multiple complexes were formed between the probe and HeLa cell extract, in the absence of competitor DNA (lane 1). One of these complexes, indicated as an arrow, was shown to be specific as it was competed by increasing concentrations (100 and 250ng) of homologous oligonucleotide (Specific) (lanes 2 and 3). Increasing concentrations (100, 250, and 500ng) of an oligonucleotide containing the immunoglobulin octamer binding element, (Octamer) (lanes 7-9), also competed for complex formation and increasing concentrations (100, 250, and 500ng) of an oligonucleotide containing a 3' G-rich sequence from the rPPT-A promoter, (G-rich) (lanes 4-6), failed to compete.


Competitor:
Specific
G-Rich
Octamer
1 2 3 4 5 6 7 8 9

-664 TTTacaTTTTGTTT -651 (reverse)

+377 TTTtgtTTTTGTTT +389

This degree of sequence homology suggests that the two elements may bind similar or related proteins. Support for this suggestion comes from the finding that increasing concentrations of the 3' T-rich oligonucleotide competes for complex formation between the 5' AT-rich element and HeLa nuclear extract (Figure 16, lanes 8-10).

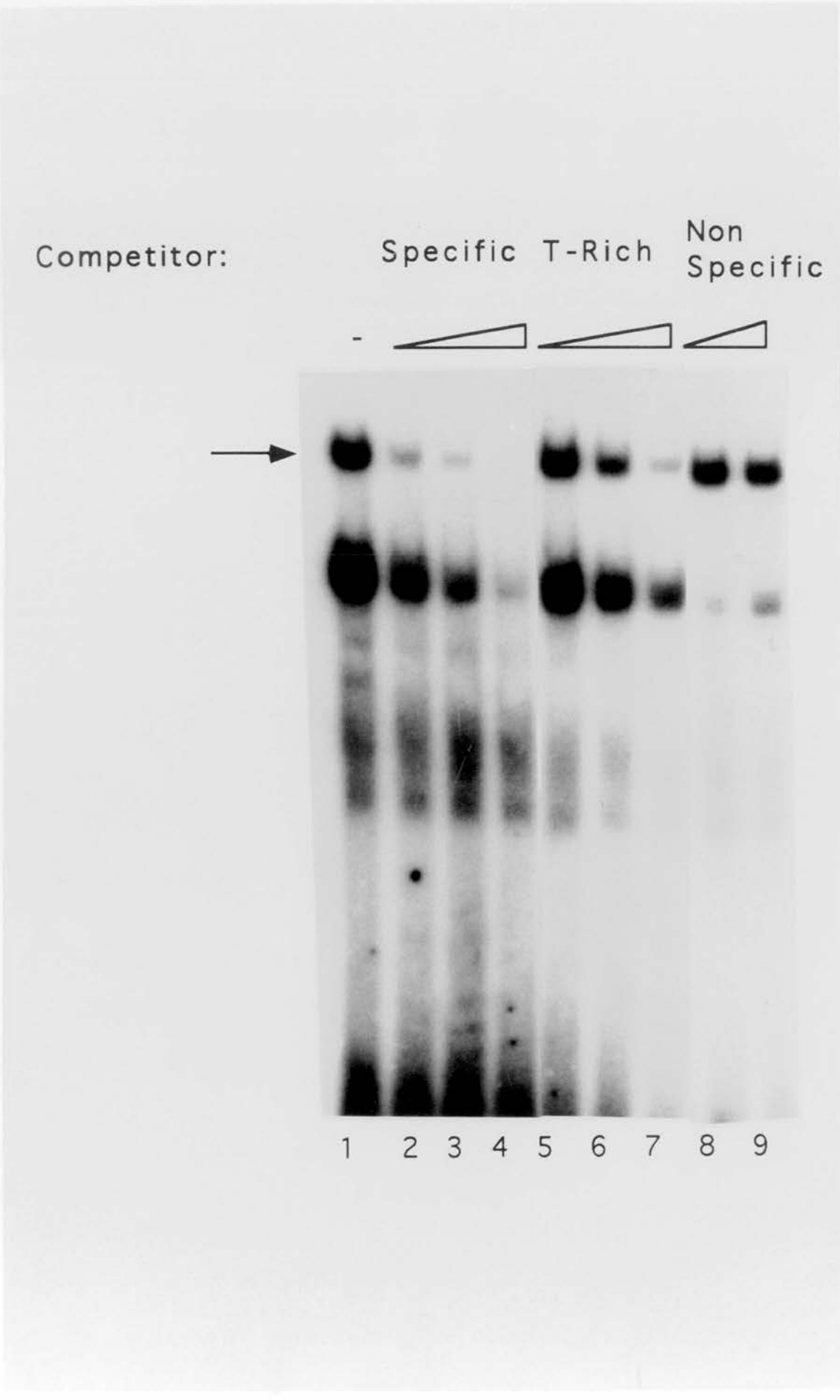
As the 5' AT-rich element has been shown to display affinity for octamer binding proteins (Section 3.6.1.11.), it was decided to determine if this is also the case for the 3' T-rich sequence. Figure 29, lanes 7-9 demonstrates that formation of a specific complex (indicated by an arrow) between the 3' T-rich oligonucleotide and HeLa nuclear extract was competed by increasing concentrations of an octamer binding element containing oligonucleotide (Oct, Figure 10). Electrophoretic mobility shift assays were also carried out using the octamer binding element containing oligonucleotide as a probe. Figure 30, lane 1 shows the position of the complex, indicated by an arrow, formed between the octamer oligonucleotide and Oct-1 from HeLa nuclear extract. The formation of this complex was competed by increasing concentrations of both homologous oligonucleotide, (Specific) (lanes 2-4), and 3' T-rich oligonucleotide, (T-rich) (lanes 5-7), but not by a non-specific oligonucleotide, (lanes 8 and 9).

Figure 17 demonstrates competition for formation of the specific complex (indicated by an arrow) between Oct-1 from HeLa cells and the octamer containing oligonucleotide by both the 5' AT-rich (lanes 5-7) and the 3' T-rich (8-10) elements. It appears from this figure that the 3' T-rich oligonucleotide has less affinity for Oct-1 than the 5' AT-rich element. Competition was complete at a 100-fold concentration excess of 5' AT-rich oligonucleotide (lane 6) but even a 200-fold concentration excess of the 3' T-rich oligonucleotide (lane 10) failed to compete the shift fully. This decreased affinity for Oct-1 complex formation by the 3' T-rich element compared to the 5' AT-rich element is not surprising based on the fact that the 3' T-rich sequence is more divergent from the octamer binding protein consensus sequence.

Figure 30. Electrophoretic mobility shift assay using the immunoglobulin octamer binding element and extract from HeLa cells.

A double stranded oligonucleotide containing the immunoglobulin octamer binding element was 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotide (1ng) was incubated with HeLa nuclear extract (20μg), in the presence of 1μg of poly d(I-C).

Multiple complexes were formed between the probe and HeLa cell extract, in the absence of competitor DNA (lane 1). One of these complexes, indicated as an arrow, was shown to be specific as it was competed by increasing concentrations (10, 25 and 100ng) of homologous oligonucleotide (Specific) (lanes 2-4). Increasing concentrations (10, 25 and 50ng) of an oligonucleotide containing a 3' T-rich sequence from the rPPT-A promoter, (T-rich) (lanes 5-7), also competed for complex formation and increasing concentrations (100 and 250ng) of a non-specific oligonucleotide (lanes 8 and 9) failed to compete.


Competitor:
Specific
T-Rich
Non Specific
-
1 2 3 4 5 6 7 8 9

3.7. Discussion.

The aim of this study was to investigate how the rPPT-A gene is regulated at the transcriptional level. To this end, this study set out to identify potential regulatory elements within a 1300 bp region of the promoter, encompassing sequences both 5' and 3' of the major transcriptional start site, by the use of exonuclease protection analysis, DNase 1 footprinting analysis and electrophoretic mobility shift assays.

3.7.1. Exonuclease protection analysis.

By the exonuclease protection assay (Quinn *et al.*, 1987), using the construct pSM1/Q5 and HeLa nuclear extract, a protein binding site was identified within the rPPT-A promoter which was estimated to lie in close proximity to the transcriptional start site (Figure 7).

The rPPT-A gene is expressed endogenously in DRG neurons and expression of constructs containing regions from the rPPT-A promoter has also been demonstrated, by microinjection, in DRG neurons (Hokfelt *et al.*, 1975; Otsuka *et al.*, 1982; Price, 1985; Tuchscherer and Seybold, 1985; Henken *et al.*, 1990; Mulderry *et al.*, 1993). However, there is a lack of established cell lines which express the gene endogenously at high levels or show significant expression of transiently transfected reporter constructs (Gilchrist *et al.*, 1991; personal communication J. Quinn). This lack of expression in tissue culture cell lines may be due to either the presence of a repressor or the absence of a specific activator in such cells. If a repressor is present, its mechanism of action may be to bind to a particular sequence within the rPPT-A promoter. As it has been shown that 5' deletion of the promoter fails to relieve repression and that individual elements identified as protein binding sites, when in isolation, will drive expression of reporter genes in transfected tissue culture cells (personal communication, J.Quinn), it is postulated that such a repressor sequence element may lie close to the transcriptional start site. The site identified by exonuclease protection analysis may therefore represent such a repressor/DNA interaction.

3.7.2. DNase 1 footprinting and electrophoretic mobility shift assays.

Exonuclease protection analysis does not allow the exact location of identified protein/DNA interaction sites to be determined. In order to achieve this, DNase 1 footprinting analysis was carried out using a series of constructs containing various fragments from the rPPT-A promoter cloned into plasmid vectors (Figure 6). Electrophoretic mobility shift assays were also carried out in order to further characterise some of the identified elements.

By the use of DNase 1 footprinting analysis and electrophoretic mobility shift assays multiple protein/DNA interaction sites have been identified within a 1300 bp region of the promoter. The nucleotide sequence of the region of the promoter analysed

is given in Figure 8 with identified protein/DNA interaction sites, numbered 1 to 19, underlined.

For the purposes of this discussion, the potential regulatory elements identified within the rPPT-A promoter have been divided into several groups based on their nature and their position within the promoter. These groups are:

a. E-Box motifs.
b. AP-1 and CRE elements.
c. Purine-rich and G-rich elements.
d. AT- and T-rich elements.
e. Unidentified elements 1 and 2.
f. Region binding single-stranded and double-stranded DNA binding proteins.
g. Minimal promoter element.
h. Region displaying differential protein binding activity.

3.7.2.1. Potential regulatory elements within the rPPT-A promoter.

3.7.2.1.1 E-box motifs.

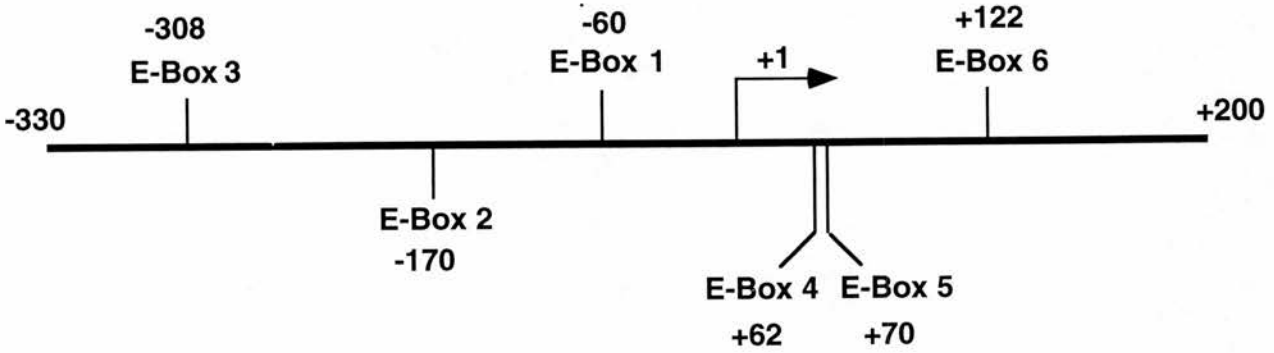
Six E-box motifs (consensus sequence CANNTG) have been identified, by DNase 1 footprinting and/or electrophoretic mobility shift assays, within a region of the rPPT-A promoter spanning approximately 450 bp of sequence. A diagrammatic representation of how these motifs are arranged on the promoter is shown in Figure 31. Three of the motifs, termed E-box 1, E-box 2 and E-box 3, are located 5' of the transcriptional start site and the other three, termed E-box 4, E-box 5 and E-box 6, are located 3' of the major transcriptional start site .

The consensus sequence CANNTG is recognised by members of the bHLH and bHLH-Zip families of transcription factors (Murre *et al.*, 1989a; Davis *et al.*, 1990; Fisher *et al.*, 1991) (Section 1.2.3.1.1.h). These proteins bind their target DNA site as dimers and can be divided into three classes (Murre *et al.*, 1989b). Class A proteins include the ubiquitously expressed bHLH proteins (termed E proteins), such as E12, E47 and E2A (Murre *et al.*, 1989b), class B proteins can form heterodimers with class A proteins and include the tissue specific bHLH proteins, such as the myogenic proteins MyoD (Davis *et al.*, 1987), Myf 5 (Braun *et al.*, 1989) and myogenin (Wright *et al.*, 1989) and class C proteins include members of the Myc protein family (DePinho *et al.*, 1987), Max proteins (Blackwood an Eisenman, 1991), Mad (for Max dimerization) (Ayer *et al.*, 1993), Mxi1 (for Max interactor 1) (Zervos *et al.*, 1993) and the transcription factors USF (Carthew *et al.*, 1985; Sawadogo and Roeder, 1985),

Figure 31. Location of identified E-box motifs within the rPPT-A promoter.

A diagrammatic representation of the rPPT-A promoter, spanning nucleotides -330 to +200 relative to the major transcriptional start site, is shown. Six identified E-box motifs (consensus sequence CANNTG), numbered E-box 1 to 6, are shown along with the position of the first nucleotide of each motif. The majority of the E-box motifs were identified by DNase 1 footprinting analysis with the exception of E-box 3 which was identified by electrophoretic mobility shift assays.

E-box 1 is contained within a footprinted element spanning nucleotides -67 to -47.
E-box 2 is contained within a footprinted element spanning nucleotides -177 to -155.
E box-3 spans nucleotides -308 to -303.
E-box 4 and E-box 5 are both contained within a footprinted element spanning nucleotides +55 to +89.
E-box 6 is contained within a footprinted element spanning nucleotides +116 to +130.


-308
E-Box 3
-60
E-Box 1
+1
+122
E-Box 6
-330
+200
E-Box 2
-170
E-Box 4
+62
E-Box 5
+70

TFE-3 (Beckmann *et al.*, 1990), TFEB (Carr and Sharp, 1990) and AP-4 (Hu *et al.*, 1990).

One cell may contain several bHLH proteins recognising the same six base pair E-box consensus sequence, therefore mechanisms must exist which allow the proteins to achieve some degree of specificity in their action. One way in which specificity of action may be achieved is by different homo- and heterodimeric protein complexes preferentially recognising different sequences both inside and outside core E-box motif (Blackwell and Weintraub, 1990; Blackwell *et al.*, 1990; Sun and Baltimore, 1991; Wright *et al.*, 1991; Fisher and Goding, 1992; Hu *et al.*, 1992; Kato *et al.*, 1992; Blackwell *et al.*, 1993; Fisher *et al.*, 1993). Interestingly, the six E-box motifs identified within the rPPT-A promoter all contain different core motifs and flanking sequences (Figure 32) suggesting that they may each bind different protein complexes.

a. E-box motifs located 5' of the transcriptional start site.

E-box 1 (element 1) corresponds to the core consensus sequence CACGTG (shown in bold type below) and is contained within the HeLa generated footprint spanning nucleotides -67 to -47 (Figures 9 A and 9B):

$^{-67}$ AGAGTGT**CACGTG**GCTCTCCA $^{-47}$

Electrophoretic mobility shift assays (Figures 11 and 12) demonstrated that proteins from various cellular and tissue extracts will form specific complexes with an oligonucleotide termed E1, containing this footprinted element and flanking sequences. These findings suggest that protein(s) present in a wide variety of tissues will potentially bind to this region of the rPPT-A promoter.

The core consensus sequence CACGTG is recognised by the class C bHLH-Zip proteins. It has subsequently been demonstrated, by electrophoretic mobility shift assays, that bacterially expressed Max, which belongs to the class C proteins, will specifically bind to the E1 oligonucleotide (J.Paterson *et al.*, submitted). Max homodimers and heterodimers, formed by association of Max with the different Myc proteins and the Max interactor proteins Mad and Mxi1, can all potentially bind to this same consensus sequence (Blackwood and Eisenman, 1991; Pendergast *et al.*, 1991; Blackwood *et al.*, 1992a and 1992b; Kato *et al.*, 1992; Littlewood *et al.*, 1992; Ayer *et al.*, 1993; Zervos *et al.*, 1993). Of all the potential dimeric complexes, Max-Max, Myc-Max, Mad-Max and Mxi1-Max, which may bind to this sequence, only Myc-Max heterodimers are thought to be capable of activating transcription (Amati *et al.*, 1992, 1993; Kato *et al.*, 1992; Kretzner *et al.*, 1992).

As Max is a constitutively, ubiquitously expressed protein (Blackwood an Eisenman, 1991) it is possible that in DRG neurons it will specifically dimerise with

		CANNTG	
E-box 1	-66	GAGTGT CACGTG GCTCTC	-49
E-box 2	-176	TTGGTC CAGATG TTATGG	-159
E-box 3	-314	TAATCT CAGGTG TCACTG	-297
E-box 4	+56	GCCCAG CAAGTG CGCACC	+73
E-box 5	+64	AGTGCG CACCTG CGGAGC	+81
E-box 6	+116	GCCTGC CATCTG CCGCTG	+133

Figure 32. Sequence comparison of six E-box motifs identified within the rPPT-A promoter.

The sequence of six E-box motifs (E-box 1 to 6) identified within the rPPT-A promoter, by DNase 1 footprinting analysis and electrophoretic mobility shift assays, plus six base pairs of 5' and 3' flanking sequence are shown.
The consensus sequence for an E-box motif is shown as CANNTG.

Myc family members or Max interactor proteins, present in such tissue, and that the resulting heterodimeric complexes will bind to the E-box 1 motif to regulate rPPT-A gene expression.

Another factor of particular interest which also recognises the core consensus sequence CACGTG is the transcription factor USF. USF was initially identified as a cellular factor which activates the Ad-ML promoter by binding to an E-box motif which, like E-box 1, is located at position -60 relative to the major transcriptional start site (Sawadago and Roeder, 1985). It has subsequently been demonstrated that a factor which shares immunological relationships with USF, termed TFII-I, binds co-operatively with USF to both the high-affinity USF site (E-box) and to two initiator elements (Inr 1 and Inr 2), distinct in sequence from the E-box motif, located near the transcriptional initiation site of the Ad-ML promoter (Roy *et al.*, 1991). Du *et al.* (1993) havesubsequently shown that USF, in addition to stimulating transcription initiation through the Ad-ML E-box site, is capable of stimulating the basal activity of both HIV-1 and Ad-ML core promoters, lacking E-box sites, through Inr elements. It was suggested from these results that many bHLH regulatory proteins may interact with both classical E-box sites and with Inr elements. As a region of sequence located between the TATA box transcriptional start site of the rPPT-A promoter has been identified as a potential regulatory element, by electrophoretic mobility shift assays and DNase 1 footprinting analysis (Section 3.6.1.13.), it is possible that bHLH factors, such as USF or TFII-I, bind to both this element and to the E-box 1 motif to regulate transcription.

E-box 2 (element 2) is contained within the HeLa generated footprinted region spanning nucleotides -177 to -155 (Figure 13) and consists of the asymmetrical consensus sequence CAGATG as shown in bold type below:

-177 TTTGGTC**CAGATG**TTATGGACTC -155

The identity of the protein(s) from HeLa cells which bind to this footprinted sequence is unknown. However, binding by proteins from HeLa cells to an oligonucleotide containing the E-box 2 motif and flanking sequences is specifically competed by related E-box containing oligonucleotides in electrophoretic mobility shift assays (personal communication, J. Quinn). This suggests, therefore, that the footprint is generated by protein(s) specifically recognising the E-box 2 motif .

The sequence CAGATG is found within the promoter regions of several peptide hormone genes, including the rat glucagon promoter where it is critical for the specific expression of glucagon in alpha cells of the pancreatic islets (Phillipe *et al.*, 1988).

DNase 1 footprinting assays using the construct pVL29 and HeLa nuclear extract (Figure 14) failed to generate any footprint covering a noted E-box consensus

sequence. This sequence, termed E-box 3 (element 5), is located immediately downstream of the AP-1' footprint (element 6), and contains the core motif CAGGTG, as shown below in bold type:

-311 TCT**CAGGTG**TCA -300

It has however been shown that an oligonucleotide containing the E-box 3 motif and flanking sequence will form a specific complex by electrophoretic mobility shift assays using HeLa cell extract. Additionally, the complex is competed by other E-box containing oligonucleotides suggesting that proteins will specifically recognise and bind to the E-box motif (personal communication, J.Quinn).

The reason for the lack of a footprint covering the E-box 3 site is unknown. Perhaps, in HeLa cells, binding to the E-box 3 site is sterically hindered by proteins which specifically recognise neighbouring AP-1 and AP-1' sites. Such interactions between factors binding to specific regulatory elements have been demonstrated for various gene promoters. For example, synergism has been reported between bHLH proteins and octamer binding proteins in the calcitonin/calcitonin gene-related peptide (CT/CGRP) promoter (Tverberg and Russo, 1993). In this study, an 18 bp sequence element, essential for cell-specific expression of the gene in rat thyroid C-cell lines, was shown to contain an overlapping E-box motif and octamer binding protein consensus sequence (GGCAGCTGTGCAAATCCT). It was subsequently demonstrated, by functional analysis and electrophoretic mobility shift assays, that both sites are synergistically involved in controlling cell-specific expression of the CT/CGRP gene (Tverberg and Russo, 1993). In another study, Yoon and Chikaraishi (1992) demonstrated the presence of a minimal enhancer element from the rat tyrosine hydroxylase promoter which is composed of an AP-1 motif and an adjacent E-box motif (TGATTCAGAGGCAGGTGCCTGTGA). It was demonstrated that an interaction between the AP-1 and E-box motifs is necessary for cell-specific tyrosine hydroxylase expression (Yoon and Chikaraishi, 1992).

Therefore, it is possible that there is an interaction between the different factors which can bind to the E-box 3 and AP-1 motifs, from the rPPT-A promoter, and that such interactions are important for regulating gene expression. It is possible that other cell or tissue types express proteins with a higher affinity for the E-box motif or that the expression of the proteins which recognise the different motifs can be altered in response to certain extracellular stimuli, thus potentially allowing for tissue-specific and stimulus-dependent gene expression.

b. E-box motifs located 3' of the transcriptional start site.

DNase 1 footprinting analysis (Figure 24) identified two sequence elements, located 3' of the major transcriptional start site, which contain three E-box motifs and are bound by proteins present in rat cerebellar nuclear extract. HeLa nuclear extract, in the same assays, also generated two footprints, however these are distinct from the cerebellar generated footprints and do not include E-box motifs. The E-box motifs contained within the cerebellar generated footprints are termed E-box 4, E-box 5 and E-box 6.

E-box 4 (CAAGTG) and E-box 5 (CACCTG) (shown below in bold type) are both contained within a footprinted element (element 13) spanning nucleotides +55 to +89:

+55 CGCCCAG**CAAGTG**CG**CACCTG**CGGAGCATCACCGG +89

and E-box 6 (CATCTG) is contained within a footprinted element (element 16) spanning nucleotides +116 to +130:

+116 GCCTGC**CATCTG**CCG +130

Although both contain E-box motifs, it is possible that the two footprints are generated by proteins recognising and binding to distinct DNA sequence motifs present within the two elements. Additionally, if the E-box motifs are specifically recognised by bHLH protein complexes it seems unlikely that both E-box 4 and E-box 5 will be bound simultaneously as they are so close to each other. The E-box 6 sequence CATCTG is present in two elements, termed Nir and Far boxes, from the rat insulin I gene promoter essential for enhancer activity (Moss *et al.*, 1988) and additionally is present in an element, termed insulin control element (ICE), from the rat insulin II gene promoter where it is bound by a positive acting factor specific to β-cells (Whelan *et al.*, 1990). The identity of the factors which do bind to these two footprinted sequences and whether they specifically recognise the E-box motifs, may be further investigated by the use of electrophoretic mobility shift assays.

As this region of the promoter appears to display differential protein binding characteristics, depending on whether the extracts used are of neuronal or non-neuronal origin, it is possible that the DNA binding proteins responsible for generating the footprints are expressed in a cell- or tissue-specific manner. There are many known transcription factors which are specifically expressed in certain cell or tissue types. An example of such proteins are the class B myogenic bHLH proteins, such MyoD (Davis *et al.*, 1987), Myf 5 (Braun *et al.*, 1989) and myogenin (Wright *et al.*, 1989), which have been shown to be specifically associated with the regulation of muscle-specific genes (Lassar *et al.*, 1991). Other bHLH proteins have been shown to be display a

neuronal-specific pattern of expression. These include the proteins encoded by genes of the *achaete-scute* complex of *Drosophila* and their rat homologues, MASH-1 and MASH-2 (mammalian achaete-scute homologues), which have been shown to play an important role in neuronal determination (Cabrera *et al.*, 1987; Johnson *et al.*, 1990, 1992; Guillemot *et al.*, 1993).

Thus, in summary six E-box motifs have been identified within a region of the rPPT-A promoter spanning approximately 450 bp. Figure 32 demonstrates that all six motifs contain different sequences both within and flanking the six base pair consensus sequence, suggesting that each may bind different combinations of bHLH proteins.

Several genes contain multiple E-box motifs within their promoter regions, including the mouse desmin gene (Li and Capetanaki, 1993) and the acetylcholine receptor δ-subunit gene (Simon and Burden, 1993), where each motif has been suggested to play a distinct role in regulating gene expression. Based on the fact that the six motifs identified within the rPPT-A promoter are predicted to preferentially bind different protein complexes and that they are located within distinct regions of the promoter, it is possible that they also have a variety of different roles to play in regulating gene expression.

The exact nature of the proteins binding to these motifs and their individual roles in regulating rPPT-A gene expression are unknown. DRG neurons have been shown to display high level of expression of the bHLH protein NSCL, cloned from a mouse cDNA library based on its homology to the hemopoietic bHLH protein SCL (Begley *et al.*, 1992). NSCL has been shown to be expressed in the developing nervous system (thus neurological SCL) where it is thought to play a role in neuronal development (Begley *et al.*, 1992). Therefore, it is possible that protein(s) similar or related to NSCL regulate rPPT-A gene expression in DRG neurons by binding to one or a combination of the identified E-box motifs within the promoter. Additionally, it has been shown that the expression of the bHLH protein MASH-1 is upregulated by NGF in PC12 cells (Johnson *et al.*, 1990, 1992). As the levels of rPPT-A mRNA have been shown to be upregulated by NGF in adult rat DRG (Lindsay and Harmar, 1989), it is tempting to postulate that one or a combination of the identified E box motifs within the rPPT-A promoter may be important for conferring the transcriptional response to NGF in DRG neurons.

3.7.2.1.2. AP-1 and CRE elements.

DNase 1 footprinting analysis identified five elements within the rPPT-A promoter which contain regions of sequence displaying homologies to an AP-1 consensus sequence (TGA G/C TCA) and/or a CRE/ATF (cyclic AMP element binding protein/activating transcription factor) binding motif (TGACGTCA).

a. AP-1 elements.

Three sequence elements, termed AP-1, AP-1' and HeLa AP-1, were identified within the rPPT-A promoter. These are all bound by proteins present in HeLa cell extract and contain regions displaying homology to an AP-1 consensus sequence (TGA G/C TCA). Additionally, an element was identified which displays homologies to both an AP-1 consensus sequence and a CRE/ATF motif. This element has been termed 'CRE' and will be discussed in Section b (CRE elements).

The consensus sequence TGA G/C TCA is recognised by the transcription factor AP-1. This factor is a dimeric complex consisting of proteins belonging to the bZIP family of DNA-binding proteins (Section 1.2.3.1.1.g.). These proteins include all of the members of the fos and jun gene families discovered to date, such as c-fos (Curran and Teich, 1982; Curran *et al.*, 1984; Curran *et al.*, 1985), c-jun (Angel *et al.*, 1987; Bohmann *et al.*, 1987; Maki *et al.*, 1987), fra-1 (Cohen and Curran, 1988; Cohen *et al.*, 1989), fra-2 (Matsui *et al.*, 1990; Nishina *et al.*, 1990; Yoshida *et al.*, 1991), fosB (Zerial *et al.*, 1989), junB (Ryder *et al.*, 1988) and junD (Hirai *et al.*, 1989; Ryder *et al.*, 1989). The proteins encoded by the jun genes can bind DNA as homodimers or can heterodimerise with Fos proteins. Therefore a large number of regulatory protein complexes can be generated among family members all of which are likely to possess different DNA-binding and transcriptional activities (Halazonetis *et al.*, 1988; Nakabeppu *et al.*, 1988; Rauscher *et al.*, 1988; Kovary and Bravo, 1991).

An element, footprinted by HeLa cell extract, was identified within the rPPT-A promoter spanning nucleotides -345 to -330 (Figure 14) (element 7), as shown below:

-345 AGCA**TGAGTCA**CTTCG -330

This footprint contains a region (shown in bold type), termed AP-1, which displays a perfect match to an AP-1 consensus sequence.

It appears, by electrophoretic mobility shift assays, that this element does represent a *bona fide* AP-1 site. Specific complexes formed between this element and extract from HeLa cells are recognised by antibodies directed against c-Fos and additionally the element will bind partially purified Fos and Jun proteins from HeLa cell extract (personal communication, J. Quinn).

An additional footprint, spanning nucleotides -324 to -308 (element 6) (Figure 14), was identified within the rPPT-A promoter which contains a region of sequence, termed AP-1', displaying a six out of seven base pair match to an AP-1 consensus sequence, as shown below in bold type:

-324 TTTTGA**TGAGTAA**TCTC -308

Although this element does not contain a perfect AP-1 consensus sequence, several studies have shown that such sites will bind AP-1 complexes. For example, Bohmann *et al.* (1987) demonstrated that the binding of purified AP-1 and c-Jun to an AP-1 recognition site from the *Saccharomyces cerevisiae his3* promoter was not affected after the introduction of certain point mutations into the recognition sequence. Additionally, a motif from the rat tyrosine hydroxylase promoter (TGATTCA), which also contains a six out of seven base pair match to an AP-1 consensus sequence, has been shown to represent a *bona fide* AP-1 motif (Yoon and Chikaraishi, 1992).

Subsequent studies have shown that the rPPT-A AP-1' sequence element, in a similar manner to the AP-1 element at -340, will bind partially purified Fos and Jun proteins from HeLa cells in electrophoretic mobility shift assays and that it will bind proteins recognised by antibodies against c-Fos (personal communication, J. Quinn). Therefore, based on these studies, this element does seem to represent an AP-1 site.

Finally, an element, footprinted using HeLa nuclear extract and located 3' of the transcriptional start site, was found to contain a region of sequence displaying a five out of seven base pair match to an AP-1 consensus sequence (Figure 24). This element (element 15), termed HeLa AP-1, spans nucleotides +98 to +107 and when nucleotide +108 is included a five out of seven base pair match (shown in bold type below) becomes apparent:

+98 GCAG**TGAGTAC** +108

It is not known if this element does represents an AP-1 site, however, as demonstrated for the AP-1' element, sites which are not perfect AP-1 consensus sequences will bind AP-1 protein complexes. Electrophoretic mobility shift assays may provide further information concerning the binding characteristics of this element and whether or not it binds members of the fos or jun gene families.

Therefore, three potential AP-1 binding sites have been identified within the rPPT-A promoter. Additionally, an element has been identified which displays homologies with both an AP-1 consensus sequence and a CRE/ATF motif, this element is discussed in more detail in Section b (CRE elements). The role that each of these sites has to play in the regulation of rPPT-A gene expression is unknown. As the fos and jun gene families have been shown to exhibit cell type, stimulus, and temporal specificity of expression (reviewed by Morgan and Curran, 1991), it is possible that these sites are involved in mediating both tissue-specific and stimulus-dependent gene expression. Of particular interest is the fact that the expression of the c-fos and c-jun genes have been shown to be induced in response to NGF in PC12 cells (Curran and Morgan, 1985; Greenberg *et al.*, 1985; Kruijer *et al.*, 1985; Milbrandt, 1986) and it has also been shown that the levels Fos, Jun and Fra proteins vary in PC12 cells in

response to NGF (Quinn, 1991). As the levels of rPPT-A mRNA have been shown to increase in adult rat DRG neurons in response to NGF treatment (Lindsay and Harmar, 1989), it is possible that this alteration in expression levels is in part mediated by factors binding to one or a combination of the identified AP-1 elements within the promoter.

b. CRE elements.

Two footprinted sequence elements, termed 'CRE' and U.D. 1, both displaying homologies to a CRE/ATF motif (TGACGTCA), have been identified within the rPPT-A promoter.

The consensus sequence TGACGTCA is recognised by proteins belonging to the CREB/ATF family such as CREB, CRE-BP1 and the ATFs (Hai *et al.*, 1989; Maekawa *et al.*, 1989; reviewed in Habener, 1990; Foulkes *et al.*, 1991). Like the members of the fos and jun gene families, these factors belong to the bZIP family of DNA-binding proteins and they also bind their DNA target sites as dimers, with multiple homo- and heterodimeric complexes forming between the different family members (Section 1.2.3.1.1.g.). CREB is believed to mediate transcriptional signals directed by the cAMP and calcium second-messenger pathways (Yamamoto *et al.*, 1988; Dash *et al.*, 1991; Sheng *et al.*, 1991).

DNase 1 footprinting analysis (Figure 13) identified an element, bound by proteins present in HeLa cells, which has previously been noted as a potential CRE from sequence analysis of the rPPT-A promoter (Chapman *et al.*, 1993). This element (shown in bold type below) is contained within a footprint (element 3) which spans nucleotides -198 to -180:

-198 AT**TGCGTCA**TTTCGAACCC -180

It can be seen that this element actually displays homologies with both an AP-1 consensus sequence (TGA G/C TCA) and with a CRE/ATF motif (TGACGTCA). However, based on the fact that this region has previously been noted as a potential CRE (Chapman *et al.*, 1993) and that the corresponding region from the bovine promoter has been shown to act as a functional CRE (Kageyama *et al.*, 1991), for the purposes of this study it is termed a 'CRE'.

Studies have shown that, when used as an oligonucleotide in electrophoretic mobility shift assays, this element will compete for complex formation between HeLa cell extract and oligonucleotides containing AP-1 and/or CRE/ATF elements (Morrison *et al.*, 1994 in press). Functional assays have also shown that this element possesses properties resembling both AP-1 and CRE elements (Morrison *et al.*, 1994 in press). Attempts to identify the proteins from HeLa cells which will interact with this element

have demonstrated that it is not bound by proteins recognised by antibodies directed against the proteins Fos, Fra-1, CREB or ATF (Morrison *et al.*, 1994 in press). However, as several CREB-related genes have been cloned (Hai *et al.*, 1989; Maekawa *et al.*, 1989; reviewed in Habener, 1990; Foulkes *et al.*, 1991) and multiple fos and jun gene families exist, it is possible that the protein(s) which do bind to this sequence element are distinct from those recognised by the Fos, Fra-1, ATF and CREB antibodies.

DNase 1 footprinting analysis (Figure 25) identified an additional element (element 18), termed unidentified element 2 (U.D. 2), which contains a region of sequence displaying homologies to a CRE/ATF motif (as shown in bold type below). This footprint is located 3' of the transcriptional start site and spans nucleotides +317 to +341:

+317 GCTCTTTG**GCACGTCA**GTAGCCTTC +341

This element was footprinted using HeLa, rat brain and rat cerebellar nuclear extracts (Figure 25), although proteins present in HeLa nuclear extract did not appear to bind as avidly as proteins present in rat brain or cerebellar extracts.

It is not known if a member of the CREB/ATF family of transcription factors will specifically bind to this sequence element. Interestingly, the footprint actually appears to consist of two halves, one extending from nucleotides +317 to +332, thus including the CRE-like region, and the other extending from nucleotides +334 to +341, with a hypersensitive site separating the two halves at nucleotide +333 (G) (underlined above). It has yet to be determined how this pattern of footprinting is generated. One possible explanation is that a CRE-like protein complex may recognise one half of the element and that other factor(s) may recognise the second half. Electrophoretic mobility shift assays should help to determine the nature of the proteins binding to this region of the promoter and to determine whether any belong to the CREB/ATF family of proteins.

Therefore, multiple potential regulatory elements have been identified within the rPPT-A promoter which may be bound by AP-1 or CREB/ATF protein complexes. The number of different dimeric protein complexes which can potentially bind to each element is very large. In addition to AP-1 and CREB/ATF proteins forming homo- and heterodimers within their own families, they can additionally heterodimerise with members of the opposite family to produce protein complexes displaying novel binding and transcriptional activities (Ivashkiv *et al.*, 1990; Jones, 1990; Hai and Curran, 1991). Yet a further degree of diversity arises from the fact that proteins belonging to the C/EBP family of bZIP proteins can bind to CRE/ATF elements, both as homodimers (Bakker and Parker, 1991; Park *et al.*, 1993) and as heterodimers formed

by association with CREB/ATF proteins (Park *et al.*, 1993; Vallejo *et al.*, 1993). Additionally, heterodimerisation has been reported between bZIP transcription factors and proteins belonging to distinct classes of DNA-binding proteins. For example, the bHLH protein MyoD has been shown to heterodimerise with the c-Jun protein (Bengal *et al.*, 1992; Li *et al.*, 1992). The fact that such a large number of protein complexes can potentially bind to the AP-1 and CRE elements identified within the rPPT-A promoter suggests that they may have a wide range of roles to play in the regulation of gene expression.

3.7.2.1.3. Purine-rich and G-rich elements.

a. Purine-rich element.

DNase 1 footprinting analysis identified an element (element 4), spanning nucleotides -284 to -264, protected from DNase 1 digestion by proteins present in HeLa nuclear extract (Figure 14). This element, as shown below, is rich in purine nucleotides on one strand with the complementary strand thereby being rich in pyrimidine nucleotides :

-284 AAGAAGAGGGGAGGGGGGCGT -264

Homopurine-homopyrimidine sequences are often found in the promoter regions of actively transcribed eukaryotic genes (Nickol and Felsenfeld, 1983; Elgin, 1984). They have been shown to be sensitive to single-stranded nucleases such as S1 (Larsen and Weintraub, 1982; Nickol and Felsenfeld, 1983; Elgin, 1984) and have been proposed to play a role in transcriptional control. Such elements are present in the promoter regions of many genes including the human U1 gene family (Htun *et al.*, 1984) and the human c-myc gene (Siebenlist *et al.*, 1984; Hay *et al.*, 1987; Davis *et al.*, 1989; Postel *et al.*, 1989) where they have been termed CT elements. CT elements have been shown to adopt an altered structure under certain conditions including high pH and negative supercoiling (Lyamichev *et al.*, 1985, 1986; Hanvey *et al.*, 1988; Htun and Dahlberg, 1988; Johnston, 1988; Kinniburgh, 1989). This unusual structure is termed H-DNA and contains both triple-stranded and single-stranded regions. It has been suggested that proteins binding to these stretches of polypurine/polypyrimidine sequences may regulate gene expression by producing a nuclease-sensitive conformation *in vivo* and so altering DNA structure.

It is thought that a large family of CT element binding proteins exist displaying overlapping affinities for different elements and possessing the ability to modulate gene expression. Factors which specifically bind to CT elements have been described by a number of groups. Such factors include NSEP-1 (nuclease-sensitive element protein-1)

(Kolluri and Kinniburgh, 1991) which has been shown, by electrophoretic mobility shift assays using various oligonucleotides, to bind both duplex CT elements and the CT-rich single strands of such elements in a sequence specific manner (Kolluri *et al.*, 1992).

Two other groups have identified factors which bind to CT elements from the human c-myc promoters. The human c-myc gene is transcriptionaly controlled, principally, by two tandem promoters P1 and P2 with the downstream promoter, P2, being predominant (Spencer and Groudine, 1990; reviewed in Marcu *et al.*, 1992). DesJardins and Hay (1993) identified and characterised components regulating transcription from the two human c-myc promoters. The regulators upstream of the P1 promoter were shown to consist of five tandem repeats of CT elements, loosely following the consensus CCTCCCCA, and the regulators upstream of the P2 promoter were shown to include a sequence element which, in the opposite orientation, resembles the CT elements of P1.

The CT elements from the *c-myc* promoters have been shown to be specifically bound by a variety of potential regulatory proteins. These include the heterogeneous ribonucleoprotein particle protein K (hnRNPK) (Takimoto *et al.*, 1993) and the Zn^{2+} finger proteins ZF87/MAZ and Sp1 (DesJardins and Hay, 1993).

Sp1 was originally identified in HeLa cells on the basis of its ability to activate the SV40 early promoter (Dyann and Tjian, 1983a). Subsequently it was shown that Sp1 recognises and binds selectively to a GC-rich sequence or GC-box, GGGCGG, present as six tandem copies in the SV40 early promoter (Dynan and Tjian, 1983b). It is now known that Sp1 can activate transcription from a variety of viral and cellular genes which have one or more GC-box recognition sequences within their 5' flanking sequences (Dynan *et al.*, 1985; Jones and Tjian, 1985; Dynan *et al.*, 1986; Geiser *et al.*, 1993; Jeang *et al.*, 1993). Following the purification of Sp1 (Briggs *et al.*, 1986) a partial cDNA encoding Sp1 was cloned from a HeLa cDNA library (Kadonaga *et al.*, 1987). It was shown that Sp1 contains three Zn^{2+} finger motifs homologous to those of the general transcription factor TFIIIA (Brown *et al.*, 1985; Miller *et al.*, 1985) and that it requires Zn^{2+} for sequence-specific binding to DNA (Kadonaga *et al.*, 1987).

It has been shown that not all GC-boxes bind Sp1 equally well and sequences outside the core hexanucleotide appear to modulate the efficiency of binding. Kadogna *et al.* (1986) and Jones *et al.* (1986) compared 36 different Sp1 binding sites and revealed a range of binding affinities differing by at least 10- to 20- fold, with individual binding sites displaying a remarkable degree of sequence variation. A consensus sequence for Sp1 binding has nevertheless been compiled as shown below:

G/T G/A GGCG G/T G/A G/A C/T

When this consensus sequence is compared with the footprinted purine-rich element, a eight out of ten base pair match (shown in bold type) is revealed:

```
                    TA     TAAT
                    GGGGCGGGGC
-284  AAGAAGA**GGGGaGGGGg**GCGT  -264
```

Interestingly, an oligonucleotide containing the footprinted CT element from the rPPT-A promoter will form specific complexes with HeLa nuclear extract, by electrophoretic mobility shift assay, and it has been demonstrated that an oligonucleotide containing a c-myc CT element will specifically compete for complex formation (personal communication, J. Quinn). This suggests that a protein(s) similar to or identical to those binding to the c-myc promoters, such as Sp1, also binds to the rPPT-A promoter to regulate gene expression

b. G-rich elements.

DNase 1 footprinting assays identified an element (element 8) rich in dG residues, and thus termed a 5' G-rich element, protected from DNase 1 digestion by proteins present in HeLa nuclear extract (Figure 15). This element spans nucleotides -575 to -546, as shown below:

```
-575  TTGGTGGGGAGGGGGGTTGGGGGGGTGTGT  -546
```

A region 3' of the transcriptional start site, also rich in dG residues (element 17), was identified by DNase 1 footprinting analysis using HeLa, rat cerebellum and rat nuclear extract (Figure 25). This element is termed a 3' G-rich element and spans nucleotides +268 to +290, as shown below:

```
+268  AGTGGGTAGGGGGCTGGGACGTT  +290
```

The protein(s) binding to this element appear, by DNase 1 footprinting analysis, to differ in their relative abundance between different tissues. Proteins present in HeLa cell extract appeared to generate a stronger footprint than those present in rat cerebellar extract and the rat brain extract did not appear to generate any footprint.

When aligned, there is a region of 92% sequence homology between the 3' and 5' G-rich elements as shown below;

```
-572  GTGGGgAGGGGG  -561

+269  GTGGGtAGGGGG  +280
```

Interestingly there is also a degree of sequence homology between the 5' and 3' G-rich elements and the purine-rich element as shown below:

-570 GGGgAGGGGGg -560

+271 GGGtAGGGGGc +281

-278 GGGgAGGGGGg -266

This homology suggests that all three elements may be bound by similar or related proteins. Oligonucleotides corresponding to all three footprinted elements will form specific complexes with proteins from HeLa nuclear extract in electrophoretic mobility shift assays (Figures 26 and 27 and personal communication, J. Quinn). Cross-competition studies, using electrophoretic mobility shift assays, may be used in order to determine whether they do indeed bind similar or related proteins.

Potential transcription factors which recognise G-rich sequences and therefore may regulate rPPT-A gene expression by binding to the 5' and 3' G-rich elements include the transcription factor Sp1. As was the case for the purine-rich element (Section a), both of the G-rich elements display an eight out of ten base pair match to the consensus sequence for Sp1, as shown below in bold type:

TA TAAT
GGGGCGGGGC

-575 TTGGT**GGGGaGGGGg**GTTGGGGGGGTGTGT -546

+268 AGTGGGTAG**GGGGCtGGGa**CGTT +290

This suggests that the protein(s) binding to both the purine- and the G-rich elements may be related to the transcription factor Sp1. Interestingly, it has been shown that when different elements within the same DNA molecule are simultaneously bound by Sp1 then the different Sp1 molecules can interact with one another by looping out the intervening DNA (Su *et al.*, 1991). Therefore, it is possible that if these three identified elements do indeed bind Sp1 then they may communicate in a similar manner. In order to determine if Sp1 will bind to any of these identified rPPT-A elements, oligonucleotides containing Sp1 sites and anti-Sp1 antisera may be used in electrophoretic mobility shift assays using the purine- or G-rich rPPT-A elements as oligonucleotide probes.

Another factor that recognises sequences rich in dG nucleotides is the HSH transcription factor AP-2 (Section 1.2.3.1.1.i.). AP-2 was first identified and purified from HeLa cells as a 50-52 KDa binding activity recognising sequences common to the enhancer elements of SV40 and the human metallothionein-II_A (hMT-II_A) genes (Haslinger and Karin, 1985; Imagawa *et al.*, 1987; Mitchell *et al.*, 1987). It was subsequently demonstrated that AP-2 is able to stimulate transcription in a binding site-

dependent manner both *in vitro* (Mitchell *et al.*, 1987) and *in vivo* (Williams *et al.*, 1988; Williams and Tjian, 1991). Binding sites for AP-2 have been found in the promoter regions of a number of cellular genes including proenkephalin (Comb *et al.*, 1986), growth hormone (Imagawa *et al.*, 1987), c-myc (Imagawa *et al.*, 1987) and acetyl-CoA carboxylase (ACC) (Park and Kim, 1993). From comparison of twelve such binding sites a consensus sequence of CCCCAGGC was determined by Mitchell *et al.* (1987), although individual sites may vary substantially from the consensus.

A cDNA encoding human AP-2 was cloned by Williams *et al* (1988) and using a series of cDNA mutants, Williams and Tjian (1991) analysed the DNA-binding and transcriptional activation properties of the AP-2 protein. By methylation interference and missing contact probe assays they determined a consensus sequence for AP-2 binding to DNA which is slightly different from the originally formulated sequence (Mitchell *et al.*, 1987): GN_4GGG.

The expression of AP-2 mRNA and protein shows a cell type specific expression pattern, being absent from the human hepatoma cell line HepG2 but present in HeLa cells (Williams *et al.*, 1988) and in DRG neurons (Mitchell *et al.*, 1991). Expression is induced by retinoic acid treatment (Williams *et al.*, 1988) and in response to adjuvant mediated inflammation (Donaldson *et al.*, 1993). The DNA-binding activity of AP-2 can be blocked by SV40 large T antigen (Mitchell *et al.*, 1987) and is stimulated by both TPA and cAMP (Imagawa *et al.*, 1987; Roesler *et al.*, 1988; Park and Kim, 1993). Thus, AP-2 can potentially act to modulate transcription from a number of genes in response to various different stimuli.

It is possible that AP-2 binds to the purine- and/or G-rich elements identified within the rPPT-A promoter as all three elements contain regions of sequence matching the consensus sequence for AP-2, GN_4GGG (Williams and Tjian, 1991), and additionally all three elements were footprinted using HeLa nuclear extract where AP-2 has been shown to be expressed (Williams *et al.*, 1988). Since AP-2 mRNA has been detected in DRG neurons (Mitchell *et al.*, 1991) and it has been shown that the levels of both AP-2 and rPPT-A mRNA increase in DRG neurons in response to adjuvant mediated inflammation *in vivo* (Donaldson *et al.*, 1992, 1993), it has been suggested that AP-2 may mediate the induction of rPPT-A gene expression in DRG neurons during the early inflammatory process (Donaldson *et al.*, 1993). Therefore, it is possible that AP-2 binds to one or a combination of the purine- or G-rich elements identified within the rPPT-A promoter to mediate the transcriptional response to induced inflammation.

Other characterised transcription factors which recognise G-rich elements include a member of the Zn^{2+} finger class of transcription factors (Section 1.2.3.1.1.e.), NGF-IA (or Egr-1 : early growth response

gene). This factor binds to the consensus sequence GCGGGGGGG and mediates growth factor and TPA activation of transcription (Chavier *et al.*, 1988; Christy *et al.*, 1988; Sukhatme *et al.*, 1988). Additionally, a protein has been described which binds to a G-rich EGF response element present in the gastrin promoter (gERE), GGGGCGGGGTGGGGGG (Merchant *et al.* (1991). Although this gERE contains an Sp1 consensus site, the specific DNA binding protein bound by the gERE in GH_4 cells was found to be distinct from previously described G-rich binding transcription factors, Sp-1, AP-2, NGF-IA and NGF-IB (Merchant *et al.*, 1991), suggesting that it is bound by a so far uncharacterised DNA-binding protein.

BGP1 (beta globin protein 1) is another factor which has been shown to bind DNA sequences rich in dG residues. The 5' flanking region of the chicken β-globin gene contains a DNase 1 hypersensitive domain spanning nucleotides -60 to -260 (Lewis *et al.*, 1988). This domain contains a (dG)- homopolymer sequence of 16 bp termed a G-string (Nickol and Felsenfeld, 1983) which, in supercoiled plasmids, can assume one or more non-B DNA conformations (Kohwi-Shigematsu and Kohwi 1985; Kohwi and Kohwi-Shigematsu 1988; Kohwi 1989). The G-string is bound by the Zn^{2+}-dependent protein BGP1 from erythroid cells (Lewis *et al.*, 1988) with the minimum binding site requiring seven dG residues (Clark *et al.*, 1990). There is a core consensus for the transcription factor Sp1 at the terminal of the G-string and although a monoclonal antibody directed against BGP1 cross-reacts weakly with Sp1 and both Sp1 and BGP1 require Zn^{2+} for their DNA binding activity, the two proteins have been shown to differ in their DNA binding specificites, chromatographic properties and molecular weights (Lewis *et al.*, 1988). The biological function of this G-string within the β-globin promoter is not clear but it has been suggested that, on binding to DNA, BGP1 may effect nucleosome positioning near the promoter (Clark *et al.* 1990). Others have suggested (Larsen and Weintraub 1982; Kohwi and Kohwi-Shigematsu 1988; Kohwi, 1989) that the interconversion of the G-string between B- and non-B-form DNA may also play an important role in the control of transcription, either directly or indirectly through alteration in nucleosome positioning or conformation (Kefalas *et al.*, 1988) .

Thus, a large number of proteins exist which specifically recognise and bind to different G-rich elements from various gene promoters to mediate both stimulus-dependent and cell-specific regulation of gene expression. This, taken together with the fact that the specific DNA recognition sequence for a given transcription factor can be highly degenerate, raises the possibility that any one of the proteins described above, or as yet unidentified proteins, may specifically recognise and bind to the rPPT-A purine- and/or G-rich sequence elements to regulate gene expression. Further studies using electrophoretic mobility shift assays, in conjunction with specific anti-sera against

candidate DNA binding proteins and oligonucleotides corresponding to the various binding elements, will enable the identity of the factor(s) binding the rPPT-A elements to be determined more accurately.

3.7.2.1.4. AT- and T-rich elements.

Two related sequence elements, one AT-rich (element 10) and located 5' of the major transcriptional start site and the other T-rich (element 19) and located 3' of the major transcriptional start site, have been identified within the rPPT-A promoter.
The 3' T-rich sequence (element 19) spans nucleotides +372 to +405, as shown below, and was identified by DNase 1 footprinting analysis using HeLa, rat brain and rat cerebellar nuclear extracts (Figure 28):

+372 TGGGTTTTT**G**TTTTTGTTTTAATCTTGGTTTAGC +405

The proteins binding to this sequence element appear, by DNase 1 footprinting analysis, to display specificity in their tissue distribution. Proteins present in rat cerebellar extract failed to generate any footprint over this region and proteins from HeLa and rat brain extracts generated differing patterns of footprinting. HeLa extract generated a footprint covering nucleotides +372 to +390 and produced a hypersensitive site at nucleotide +381 (dG nucleotide, shown in bold type above). The brain generated footprint is extended further in a 3' direction, to nucleotide +405, and does not contain any hypersensitive sites.

The 5' AT-rich sequence (element 10) spans nucleotides -662 to -655, as shown below in bold type:

-665 TAA**ATGTAAAA**CAAA -651

This element is located in a region of the rPPT-A promoter not cleaved by DNase 1 even in the absence of protein extract. However, it was noted that it displays a six out of eight base pair match to the octamer binding protein consensus sequence (ATGCAAT). Subsequently, it was demonstrated, by electrophoretic mobility shift assays, that an oligonucleotide probe containing the 5' AT-rich element and flanking sequences will form sequence specific complexes with proteins present in HeLa and C1300 cells (Figures 16-18). Interestingly, these specific complexes were shown to migrate to a similar position as a complex formed between Oct-1 and an octamer containing oligonucleotide (Figure 16) and, additionally, the 5' AT-rich and octamer oligonucleotides were found to cross-compete for specific complex formation (Figure 16-18). These results then suggest that the 5' AT-rich element may bind a member(s) of the octamer binding protein family.

The 5' AT-rich sequence element was noted as sharing a significant degree of sequence homology (79%) with the 3' T-rich element, on the complementary strand.

Consequently oligonucleotides containing the two elements were found to cross-compete for complex formation in electrophoretic mobility shift assays (Figure 16). This suggests that the two elements may bind similar or related proteins. Consistent with this sequence homology and cross-competition, the 3' T-rich element was also found to display affinity for octamer binding proteins. Formation of a specific complex between the 3' T-rich oligonucleotide and HeLa nuclear extract was competed by increasing concentrations of the octamer oligonucleotide (Figure 29). Additionally the 3' T-rich oligonucleotide was shown to compete for complex formation between the octamer containing oligonucleotide and Oct-1 present in HeLa nuclear extract (Figure 17 and 30). Competition for Oct-1 complex formation required a higher concentration of the 3' T-rich oligonucleotide than was required for competition using the 5' AT-rich oligonucleotide (Figure 17). This may be expected based on the fact that the 3' T-rich sequence displays less sequence homology with the octamer binding protein consensus sequence than the 5' AT-rich sequence does.

Therefore, it is possible that both the 5' AT-rich and the 3' T-rich rPPT-A elements regulate rPPT-A gene expression by binding member(s) of the octamer binding protein family (Section 1.2.3.1.1.d.). The identity of the particular octamer binding proteins which interact with these elements remains unknown. However, the fact that the 3' T-rich element displays tissue specificity in the proteins it binds, by DNase 1 footprinting analysis, suggests that ubiquitous Oct-1 is not the only factor interacting with this region of the promoter. Different octamer binding proteins have been shown to display varying and overlapping tissue distributions and additionally the expression of many has been shown to be regulated throughout development and in response to various extracellular stimuli (He *et al.*, 1989). Therefore, the particular set of octamer binding proteins present in a given cell at any one time and in response to extracellular stimuli will vary, thus potentially allowing for tissue-specific and stimulus-dependent gene regulation.

Although the rPPT-A gene is expressed in a few non-neuronal cells, including the intestine, stomach, thyroid and retina, its expression is mainly restricted to specific populations of neuronal cells of the central and peripheral nervous systems and in particular to a subset of primary sensory neurons of the DRG (Nawa *et al.*, 1983, 1984; Krause *et al.*, 1987, 1989; Carter and Krause, 1990; Harmar *et al.*, 1990; Bannon *et al.*, 1992). Adult rat DRG sensory neurons are known to express a number of POU-domain proteins, including Oct-1, N-Oct-2, Brn-3 and a number of Brn-3-related factors (Wood *et al.*, 1992). Therefore one of these proteins may be a candidate for regulating rPPT-A gene expression in sensory neurons via the identified octamer-like elements. Interestingly, in DRG sensory neurons the levels of N-Oct-2 and its mRNA were found to be upregulated by NGF treatment (Wood *et al.*, 1992). As the

levels of rPPT-A transcripts in adult DRG neurons have been shown to both depend on NGF availability (Henken *et al.*, 1990) and to increase with NGF treatment (Lindsay and Harmar, 1989), it is tempting to postulate that N-Oct 2 may bind to one or both of the identified octamer-like elements within the rPPT-A promoter to regulate gene expression and to mediate the transcriptional response to NGF.

In order to determine the identity of the octamer binding proteins which will bind to the identified rPPT-A elements, electrophoretic mobility shift assays may be carried out using either purified octamer binding proteins or specific anti-sera directed against the individual POU-domain proteins. In order to determine if the elements are involved in mediating the transcriptional response to NGF various functional assays may be carried out, including the transfection of plasmids containing the elements linked to reporter genes into clonal cell lines and DRG neurons in culture followed by treatment with NGF.

3.7.2.1.5. Unidentified elements 1 and 2.

DNase 1 footprinting analysis identified two potential regulatory sequence elements, U.D. 1 (element 9) and U.D. 2 (element 18), located 5' and 3' of the rPPT-A promoter transcriptional start site respectively.

U.D. 1 was identified by DNase 1 footprinting assays using HeLa nuclear extract (Figure 15) and spans nucleotides -618 to -607, as shown below:

$^{-618}$ AACACTTTTGCA $^{-607}$

U.D. 2 was identified by DNase 1 footprinting assays using HeLa, rat brain and rat cerebellar nuclear extracts (Figure 25) and spans nucleotides +317 to +341, as shown below:

$^{+317}$ GCTCTTTGGCACGTCAGTAGCCTTC $^{+341}$

The proteins binding to this element appear, by DNase 1 footprinting analysis, to vary in their abundance between different tissues as proteins present in HeLa nuclear extract generated a weaker footprint than those present in rat total brain or cerebellar extracts.

When aligned, the U.D. 1 and U.D. 2 elements display an 80% sequence homology, as shown below:

$^{-616}$ CaCTTTtGCA $^{-607}$

$^{+318}$ CtCTTTgGCA $^{+327}$

In addition to this high degree of sequence homology, the two elements are placed a similar context within the promoter i.e. between a G-rich element and a T- or AT-rich element (Figure 37). Therefore, it is possible that the two elements may bind

similar or related proteins. Electrophoretic mobility shift assays will be required to be carried out in order to further investigate the potential relationship between the two elements.

As mentioned in Section 3.7.2.1.2.b. the U.D. 2 element contains a region of sequence which displays a six out of eight base pair match to a CRE/ATF motif, as shown in bold type below:

$^{+317}$ GCTCTTTG**GCACGTCA**G**T**AGCCTTC $^{+341}$

Interestingly, the footprint appears to actually consist of two halves, with the CRE/ATF-like motif being contained within the first half and being separated from the second half by a hypersensitive site corresponding to nucleotide +333 (G) (underlined above).

Unlike U.D. 2, sequences contained within the U.D. 1 element do not appear to display any homologies with a CRE/ATF motif. Therefore, if the two elements do bind similar or related proteins it seems unlikely that they will belong to the CREB/ATF family of transcription factors.

Further studies are required in order to determine the exact nature of the proteins binding to these two elements and how they are related to each other. These studies may include electrophoretic mobility shift assays using the two elements as probes in conjunction with various competitor oligonucleotides containing the consensus sequence for known transcription factors and specific anti-sera directed against various known transcription factors.

3.7.2.1.6. Region binding single-stranded and double-stranded DNA binding proteins.

DNase 1 footprinting analysis, using HeLa nuclear extract, detected protein binding activity (element 11) in a region of the rPPT-A promoter previously reported to bind a sequence-specific single-stranded DNA binding protein (Quinn and McAllister, 1993).

In their study, Quinn and McAllister postulated that a region from the rPPT-A promoter, spanning nucleotides -761 to -741, may play a role in the regulation of gene expression. This stemmed from the fact that this region of the promoter shares a significant degree of homology with an element from the rat type II Na^{2+} channel promoter, thought to be important for mediating cell-specific gene expression (Maue *et al.*, 1990). By electrophoretic mobility shift assays, it was demonstrated that proteins present in rat DRG, spinal cord, caudate and spleen extracts bound specifically to a single-stranded form of an oligonucleotide containing sequence from this region of the rPPT-A promoter (Figure 33, SS Element). Binding, however, was not detected by

Figure 33. Nucleotide sequence of a region of the rPPT-A promoter which is proposed to form a stem-loop structure, binds single-stranded DNA binding proteins and is footprinted using HeLa nuclear extract.

The sequences of a region of the rPPT-A promoter proposed to form a stem-loop structure and the loop of such a structure are shown.
Shown below these is the sequence of a footprint (footprint 1) generated by HeLa nuclear extract within this region of the promoter. Footprint 2 was generated when an oligonucleotide corresponding to the stem-loop structure was used as a competitor in DNase 1 footprinting assays.
The sequence of the region of the rPPT-A promoter demonstrated to specifically bind single-stranded DNA binding proteins (SS element) is also shown.

Stem-Loop: -779 CCCTTCTGCCTTCAGGGTGTGCCTGGGAAGAAGCTGTAGGGG -738

Loop: -766 AGGGTGTGCCTGGGAA -751

Footprint 1: -778 CCCTTCTGCCTTCAGGGTGTGCCTG -755

Footprint 2: -759 GCCTGGGAAGAAGCTGTAGGGGAACAAAA -731

SS Element: -765 GGGTGTGCCTGGGAAGAAGCTGTAGGG -752

proteins present in rat testis extract, HeLa nuclear extract and a number of other clonal cell lines (Quinn and McAllister, 1993).

The DNA sequence of this element was subsequently shown to share homologies with other single-stranded elements from various promoters including an element, termed M3, from the adipsin gene (Wilkison *et al.*, 1990) and a proximal repressor element (PRE) from the growth hormone promoter (Pan *et al.*, 1990). This led to the suggestion that the protein(s) which specifically recognise and bind to the single-stranded form of this DNA element may act to regulate rPPT-A gene transcription (Quinn and McAllister, 1993).

Surprisingly, a significant degree of homology was also revealed between the rPPT-A single-stranded DNA element and an RNA stem-loop-structure, termed TAR element, located at the 5' end of all HIV-1 transcripts (Rosen *et al.*, 1985). This element is bound by the potent activator of viral gene transcription, Tat (Ayra *et al.*, 1985; Sodroski *et al.*, 1985). TAR has been shown to be capable of forming a stem-loop structure (Muesing *et al.*, 1987) and this structure has been shown to be crucial for transactivation by Tat (Feng and Holland, 1988). Quinn and McAllister (1993) suggested that a putative stem-loop structure may also form within the rPPT-A promoter, with the sequences shown to bind a single-stranded DNA binding protein forming the loop of such a putative structure. It was suggested if such a stem-loop structure does occur *in vivo* then proteins with high affinity for the single-stranded form of the DNA would bind preferentially to the loop and perhaps act to stabilise the structure.

Subsequent electrophoretic mobility shift assays have been carried out using an oligonucleotide which spans nucleotides -790 to -738 and so includes the entire putative stem-loop structure, from nucleotides -779 to -738 (Figure 33, Stem-Loop) (personal communication, J.Quinn). These assays reveal a further degree of complexity in the nature of protein/DNA interactions within this region of the promoter. As was found for the shorter oligonucleotide, proteins from rat tissue extracts were found to preferentially bind to a single-stranded form of the oligonucleotide for which proteins from HeLa nuclear extract and a number of other cell lines displayed no affinity. However, in contrast to the shorter oligonucleotide, the presence of the additional nucleotides appeared to allow proteins from HeLa nuclear extract and a number of other cell lines and tissues to bind specifically to the double-stranded form of the oligonucleotide (personal communication, J. Quinn).

In order to further investigate the binding of proteins from HeLa cells to this region of the promoter, DNase 1 footprinting analysis was carried out (Figure 19). A footprint (Figure 19, Footprint 1) was generated by proteins present in HeLa nuclear extract spanning nucleotides -779 to -755, as shown below;

$^{-779}$ CCCTTCTGCCTTCAGGGTGTGCCTG $^{-755}$

Preincubation of HeLa nuclear extract with the oligonucleotide spanning the entire putative stem-loop structure appeared to compete off the complex binding to this site and to allow the binding of another factor to a distinct yet overlapping sequence (corresponding to Footprint 2, Figure 19), as shown below:

$^{-759}$ GCCTGGGAAGAAGCTGTAGGGGAACAAAA $^{-731}$

When these two footprinted regions are aligned with the proposed stem-loop sequence (Figure 33), it can be seen that footprint 1 appears to cover the 5' end of the structure and footprint 2 appears to cover the 3' end and adjacent sequence. This seems to indicate that at least two complexes can bind to this region of the rPPT-A promoter and that the binding of one appears to prevent the binding of the other. Interestingly, most of the sequence protected in footprint 2 is contained within the competitor stem-loop oligonucleotide suggesting that the binding of protein(s) element may be mediated by nucleotides -737 to -731.

It can be seen from Figure 33 that both footprint 1 and footprint 2 contain regions of sequence contained within the element shown to bind single-stranded DNA binding proteins. Thus, it appears that distinct yet overlapping regions from the rPPT-A promoter can bind both single-stranded and double-stranded DNA binding proteins.

Although the majority of reported eukaryotic sequence-specific DNA binding proteins recognise native double-stranded sequences, there have been reports of proteins which recognise the single-stranded form of DNA and are involved in the regulation of transcription. Such proteins include muscle factor 3 (MF3) which interacts with the single-stranded forms of DNA sequence elements important for regulating muscle gene expression (Santoro *et al.*, 1991) and the estrogen receptor (ER) which, by way of a DNA binding stimulating factor (DBSF), preferentially binds to the single-stranded form of its response element (Lannigan and Notides, 1989; Mukherjee and Chambon, 1990). Other factors include the sterol-regulatory element binding protein (SRE-BP) (Stark *et al.*, 1992). SRE-BP can bind specifically to both double- and single-stranded forms of its response element in the low density lipoprotein (LDL) receptor promoter with the sequence specificites for single-and double-stranded binding shown, by electrophoretic mobility shift assay, to be distinct yet overlapping (Stark *et al.*, 1992; Yokoyama *et al.*, 1993). Two distinct factors, termed PRE-binding nuclear protein (PREB) and single-stranded PREB (ssPREB), bind the double- and the single-stranded forms of the growth hormone promoter repressor element (PRE) respectively (Pan *et al.*, 1990). These two factors have been shown to display distinct patterns of expression (Pan *et al.*, 1990). PREB has been shown, by electrophoretic mobility shift

assays, to be specifically expressed in non-pituitary cell-lines while ssPRE was detected in all cell lines tested (Pan *et al.*, 1990).

Regulatory sequence elements which can adopt a single-stranded structure include CT elements (Section 3.7.2.1.3.a.). The rPPT-A footprinted sequence, footprint 1, does contain a high degree of dC and dT nucleotides, particularly at its 5' end, and footprint 2 contains a high degree of dG and dA nucleotides. Therefore, it is possible that factors which have been shown to specifically recognise and bind to such elements, including Sp1, may bind to this region of the rPPT-A promoter to regulate gene expression.

Therefore, it appears that a wide range of proteins exist with the ability to recognise and bind to the single-stranded forms of DNA sequence elements. Additionally, many of these elements are also recognised in the double-stranded form, both by the same and distinct proteins which often display distinct yet overlapping sequence specificity and patterns of expression.

The exact nature of the proteins binding to both the single- and double-stranded forms of DNA from the rPPT-A promoter and their roles in regulating gene expression are yet to be determined. It is possible that the abundance of such proteins may differ between different cell types and in response to extracellular stimuli, thus potentially allowing for tissue-specific and stimulus-dependent gene regulation. For example, protein(s) specific to a particular set of tissues (e.g. DRG neurons) may recognise and bind to the single-stranded form of the putative stem-loop structure and other proteins, expressed in a distinct set of tissues (e.g. spleen), may preferentially recognise the double-stranded form thus imposing a different form of regulation on the promoter.

Electrophoretic mobility shift assays may be carried out in order to identify the factors which can bind to this region of the rPPT-A promoter and to determine whether they recognise the double-stranded, single-stranded or both forms of the DNA sequence

3.7.2.1.7. Minimal promoter element.

A region of the rPPT-A promoter, located between the TATA box and the major transcriptional start site, termed the minimal promoter element was noted as sharing a significant degree of sequence homology with a similarly placed element from the pro-opiomelanocortin (POMC) promoter (Section 3.6.1.13.). The POMC element has been shown to be bound by a novel trans-acting factor, termed PO-B (Riegel *et al.*, 1990). Although the binding site for PO-B is located close to the binding site for the general transcription factor TFIID (Section 1.2.2.2.a.), it has been shown that both factors are unrelated in their binding, fractionation and transcriptional stimulatory properties (Riegel *et al.*, 1990).

PO-B has been shown to be required for maximal efficiency of the POMC promoter both *in vivo* and *in vitro* and in association with the functional TATA box is capable of directing transcriptional initiation (Riegel *et al.*, 1990). Interestingly, it has been shown that factors with PO-B like binding activity are present in a number of cell lines in which POMC is not normally expressed (Reigel *et al.*, 1990), suggesting that PO-B may have a function in the regulation of other eukaryotic genes. This finding taken together with the fact that rPPT-A promoter sequence, between the TATA box and transcriptional start site, shares a significant degree of homology with the PO-B binding region suggested that PO-B, or a related factor, may bind to the rPPT-A minimal promoter element.

In order to investigate any possible relationships between the factors which bind to the POMC and rPPT-A promoters, electrophoretic mobility shift assays were carried out using oligonucleotides containing sequences either from the rPPT-A promoter, spanning nucleotides -20 to +4 (termed 168/169), or from the POMC promoter, spanning nucleotides -23 to +5. The results from such assays demonstrated that although protein(s) will specifically bind to the rPPT-A minimal promoter element (Figure 20), they do not appear to be related to PO-B. Specific complexes formed between the two oligonucleotides and various cellular extracts were shown to display very different mobilities (Figure 22) and the two oligonucleotides failed to show any cross-competition for complex formation with HeLa nuclear extract (Figure 21).

DNase 1 footprinting analysis using HeLa nuclear extract subsequently identified a protein binding site (element 12) located between the TATA box and transcriptional start site of the rPPT-A promoter (Figure 23). A footprint was generated, as shown below in bold type, which covers two dA nucleotides at positions -16 and -17 and a hypersensitive site was generated which corresponds to a dC nucleotide at position -19:

-20 GCG**AA**GC -14

This footprint is contained within the sequence of the oligonucleotide (168/169) used for the electrophoretic mobility shift assays described above. Therefore the results from both assays, taken together, demonstrate the presence of a potential regulatory element within the rPPT-A promoter, spanning nucleotides -20 and +4, as shown below:

-20 GCGAAGCAGGAGCAGGGACTAGAG +4

The identity of the protein(s) which bind to the rPPT-A minimal promoter element have yet to be determined. The core promoters of genes transcribed by RNA polymerase II commonly contain a TATA element (Section 1.2.2.) and/or an initiator (Inr) element (Grosschedl and Birnstiel, 1980; Smale and Baltimore, 1989) (Section

1.2.2.4.). Both of these elements are important for promoter function and specifically interact with sequence-specific transcription factors. Directly acting transcriptional repressor proteins have also been shown to bind to the minimal promoter regions of several genes. Thus, it is possible that a factor(s) which specifically recognises TATA, Inr or repressor elements is responsible for generating the observed binding activity over the rPPT-A minimal promoter element.

a. General transcription factors.

It is unlikely that the TATA-box binding protein, TBP (Section 1.2.2.2.a.), is responsible for generating the observed binding activity, as the TATA box from the rPPT-A promoter (identified by Carter and Krause, 1990) spans nucleotides -22 to -28 and therefore lies immediately 3' of the identified protein binding site. Other general transcription factors (Section 1.2.2.2.) are also unlikely to be responsible as none have been shown to display sequence-specific binding to DNA in the vicinity of the transcriptional start site.

b. Initiator elements.

Inr elements (Section 1.2.2.4.) are found in the vicinity of the transcriptional start site of several eukaryotic genes. Multiple types of Inr elements have been shown exist, serving a variety functions and interacting with a wide range of factors including USF and TFII-I (Roy *et al.*, 1991; Du et al., 1993), YY1 (Seto *et al.*, 1991; Shi *et al.*, 1991) and the Inr-binding protein of SV40 (IBP-s) (Wiley *et al.*, 1993).

A weak consensus sequence, as shown below, has been derived for Inr elements (Corden *et al.*, 1980) :

5'-PyPyCAPyPyPyPyPy-3'

When the sequence of the rPPT-A minimal promoter element is shown on the complementary strand, two overlapping regions (shown below in bold type) are revealed which show a 90% degree of sequence homology with this consensus sequence:

-20 CGC**TTCgTCCTC**GTCCCTGATCTC +4

-20 CGCTTCGTC**CTCgTCCCT**GATCTC +4

It is not known if the rPPT-A minimal promoter element represents such an Inr element. However, it does posses a high degree of sequence similarity with the consensus sequence and it is interesting to note that both the AdML and the rPPT-A promoters posses similarly located E-box motifs (i.e. at positions -60) and Inr-like elements close to their transcriptional start sites (Sawadago and Roeder, 1985; Roy *et*

al., 1991; Du *et al.*, 1993) (Section 3.7.2.1.1.a.). The sequence of the Inr element from the AdML promoter located close to its transcriptional start site is shown below:

$^{-3}$ CTCACTCTCTTC $^{+9}$

When this is aligned with the minimal promoter element from the rPPT-A promoter a degree of homology (62 %) is revealed:

$^{-3}$ CTCa-CtCTctTC $^{+9}$

$^{-20}$ CGCTTCGTCCTCgtCcCTgaTCTC $^{+4}$

The AdML Inr elements are bound by the same bHLH proteins which bind to its E-box motif at -60, raising the possibility that similar factors may bind to the E-box 1 motif and the minimal promoter element from the rPPT-A promoter. Further studies are required in order to determine the nature of the protein(s) binding to the rPPT-A minimal promoter element, to determine if they are related to other Inr or E-box binding proteins and to determine if they share any functional properties with other well characterised Inr elements.

c. Transcriptional repressor proteins.

It is possible that, as was found for the PO-B binding region of the POMC promoter (Riegel *et al.*, 1990), a factor other than an Inr element binding protein binds to rPPT-A minimal promoter region.

Some of the factors which are involved in mediating the repression of transcription initiation bind to sequences around the transcriptional start site. Directly acting repressors which alter the interaction of RNA polymerase II or the general transcription factors with promoters have been commonly observed in prokaryotic gene regulation (reviewed by Adhya 1989). This mechanism is also thought to account for the autoregulation of certain viruses such as regulation of SV40 gene expression by the SV40 T antigen (Rio and Tjian, 1983). Gomez-Cuadrado *et al.* (1992) reported the isolation of a DNA-binding protein from mature chicken erythrocytes, termed cIBR (chicken Inr binding protein). cIBR recognises sequences spanning the transcriptional start site of the histone H5 gene but was shown to be functionally unrelated to Inr binding proteins (Gomez-Cuadrado *et al.*, 1992). It was demonstrated that the binding of cIBR could repress transcription from the H5 promoter *in vitro* and this was thought to be achieved through direct interference with general transcription factors, other than TFIID (Gomez-Cuadrado *et al.*, 1992).

Other identified transcriptional repressor proteins proposed to mediate their action through direct communication with general transcription factors, at minimal promoter elements, include the p53 gene product. The p53 protein has been shown to act as an activator of gene expression through the binding of specific regulatory

sequences termed p53 response elements (Farmer *et al.*, 1992; Zambetti *et al.*, 1992). p53 has also been shown to negatively regulate a number of genes that lack a p53 response element including the c-fos and c-jun genes and the p53 gene itself (Ginsberg *et al.*, 1991; Santhanam *et al.*, 1991). Seto *et al.* (1992) demonstrated that p53 can inhibit transcription from minimal promoters in a HeLa nuclear extract and that it can bind to human TFIID. From these findings it was suggested that p53 mediates transcriptional repression by binding to TFIID and interfering with transcriptional initiation. Mack *et al.* (1993) subsequently demonstrated that p53 specifically represses the activity of promoters whose initiation is dependent on the presence of a TATA box. The repression observed was again consistent with a direct interaction of p53 with the TATA-dependent basal machinery.

Thus, multiple different factors, including TFIID, Inr binding factors and repressor proteins, have been shown to bind sequences around the TATA box and transcriptional start site of genes transcribed by RNA polymerase II. The identity and the function of the factor(s) interacting with the rPPT-A minimal promoter element is as yet unknown. However, it would be attractive to postulate that the factor(s) binding to this region may in some way interact with the basal transcriptional machinery to repress transcription. This would be consistent with results demonstrating that the rPPT-A promoter is not expressed, either endogenously or by transfection, in tissue culture cells and that 5' deletion of the promoter fails to relieve this repression (J. Quinn, personal communication). If a factor(s) is present in such cells which can bind to the minimal promoter element and prevent the initiation of transcription by the basal transcriptional machinery, then a lack of gene expression in these cells would be expected. If this hypothesis is correct then one would expect that cells or tissues which do express the rPPT-A gene, by transfection or endogenously, either do not express the repressor or express it in a modified form.

Further studies are required in order to identify both the nature and the expression pattern of the factor(s) which interact with this region of the promoter. Functional studies will then allow the validity of the above hypothesis to be tested.

3.7.2.1.8. Region displaying differential protein binding activity.

By DNase 1 footprinting analysis (Figure 24) extracts form rat cerebellar tissue and HeLa cells have been shown to protect distinct yet overlapping sequences from the rPPT-A promoter, in a region located 3' of the major transcriptional start site, spanning nucleotides +55 to +130.

Proteins present in HeLa nuclear extracted generated two footprints in this region, one is located in exon 1 (element 14) and spans nucleotides +74 to +94:

+74 TGCGGAGCATCACCGGGTCCG +94

and the other is located over the intron 1/exon 1 boundary (element 15) and spans nucleotides +98 to +107:

+98 GCAG**TGAGTA** +107

The identity of the factors from HeLa cells which bind to these two sequence elements are unknown. However, a region of sequence contained within the footprint covering the intron 1/exon 1 boundary, when extended to nucleotide +108 (C), displays a five out of seven base pair match to an AP-1 consensus sequence (TGA G/C TCA), as shown in bold type above. This then raises the possibility that an AP-1 complex may bind to this site to regulate transcription (Section 3.7.2.1.2.a.). Electrophoretic mobility shift assays using the two footprinted elements as oligonucleotide probes and various competitor oligonucleotides may provide further information concerning their binding characteristics.

Rat cerebellar extract also generated two footprints over this region of the promoter. One of the sequences bound by cerebellar proteins, contained within exon 1, overlaps one of the elements (element 14) bound by HeLa extract at its 5' end, but is extended further in a 3' direction. This element (element 13) spans nucleotides +55 to +88, with a hypersensitive site present at nucleotide +89:

+55 CGCCCAG**CAAGTG**CG**CACCTG**CGGAGCATCACCGG +89

The second cerebellar generated footprint (element 16) is distinct from either of the HeLa generated footprints and is contained within intron 1, spanning nucleotides +116 to +130:

+116 GCCTGC**CATCTG**CCG +130

Interestingly, both of the cerebellar generated footprints contain E-box motifs, shown in bold type (Section 3.7.2.1.1.b.). Electrophoretic mobility shift assays may help to determine the nature of the proteins binding to these footprints and to determine whether proteins specifically recognise the E-box motifs or if other sequences within the footprints are specifically recognised by distinct proteins.

A neuronal-specific protein/DNA interaction, within the rPPT-A promoter, has previously been reported by use of the exonuclease protection assay (Quinn, 1992). It was found that proteins present in rat cerebellar, hippocampal and spinal cord extracts and in extract from PC12 cells generated a specific protein/DNA interaction which was not observed by proteins present in rat spleen extract or the non-neuronal derived cell lines HeLa and 293 (Quinn, 1992). This apparently neuronal specific interaction was mapped to a region 3' of the transcriptional start site, approximately at the intron 1/exon 1 boundary (Quinn, 1992). The DNase 1 footprinting data above is consistent with this finding of a 3' neuronal specific protein/DNA interaction, as cerebellar extract was

found to generate a distinct pattern of binding from that generated by HeLa extract. Additionally, DNase 1 footprinting has allowed the sites of protein/DNA interactions to be mapped more accurately. However, where as exonuclease protection analysis suggested a lack of binding by proteins present in HeLa cell extract to this 3' region (Quinn, 1992), DNase 1 footprinting analysis has demonstrated that two sequence elements are bound by HeLa cell extract. The reasons for the discrepancy between the two different assays are unknown. However, they are very different assays and employ different approaches in the identification of protein/DNA interactions. Electrophoretic mobility shift assays using the identified elements as oligonucleotide probes and extracts of both neuronal and non-neuronal origin may provide further information regarding the binding characteristics of proteins to this region of the promoter.

The results from this study therefore suggest that a 3' region of the rPPT-A promoter, spanning nucleotides +55 to +130, displays tissue-specificity in its protein/DNA interactions. A hypothetical model to explain the observed binding pattern supposes that the expression of the transcription factors which recognise the regulatory elements, identified within this region of the rPPT-A promoter, varies between HeLa and rat cerebellar extracts or that the factors which bind to these elements are expressed in a tissue-specific manner. In this way rat cerebellar extract may contain factor(s) which preferentially recognise a particular motif or set of motifs and HeLa cells may express a different set of factors which recognise a distinct set of motifs.

Further studies, including electrophoretic mobility shift assays, should help to further characterise the binding of proteins from different cell and tissue extracts to this region of the promoter.

3.7.3. Arrangement of potential regulatory elements within the rPPT-A promoter.

Multiple potential regulatory elements have been identified within a 1300 bp region of the rPPT-A promoter, spanning nucleotides -865 to +447.

A diagrammatic representation of how these elements are arranged within the rPPT-A promoter is shown in Figure 34.

From this figure it appears that several domains, consisting of related elements, can be defined within the promoter. These include a region located 5' of the transcriptional start site, spanning nucleotides -40 to -360bp, termed the proximal region, and a flanking region consisting of the AT-, T-, G-rich and U.D. elements, located both 5' and 3' of the transcriptional start site.

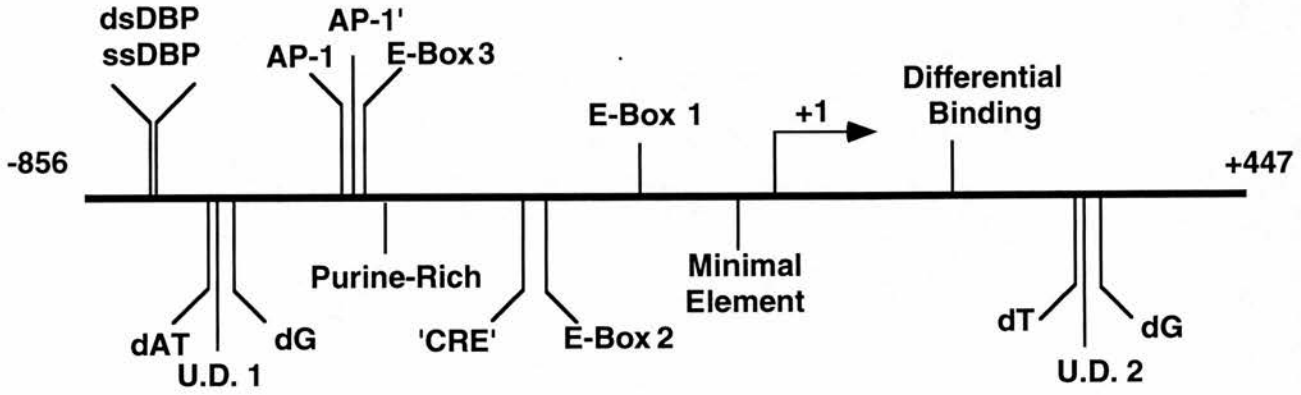
3.7.3.1. Proximal region.

Multiple putative transcription factor binding sites have been identified within a region of the rPPT-A promoter, spanning 320bp, located 5' of the transcriptional start

Figure 34. Arrangement of identified potential regulatory elements within a 1300 bp region of the rPPT-A promoter.

A diagrammatic representation of the rPPT-A promoter with identified protein/DNA interaction sites is shown. Most of these sites were characterised by DNase 1 footprinting analysis and further characterised by electrophoretic mobility shift assays.

dsDBP/ssDBP	Regions which binds both single-stranded (ss) and double-stranded (ds) DNA binding proteins (DBP). Span nucleotides -778 to -731.
dAT	Region rich in dA and dT nucleotides containing an octamer-like consensus sequence (identified by electrophoretic mobility shift assays). Spans nucleotides -662 to -655.
U.D. 1	Unidentified Element 1. Spans nucleotides -618 to -607.
dG	Regions rich in dG nucleotides. Span nucleotides -575 to -546.
AP-1	Region containing a perfect AP-1 consensus sequence. Spans nucleotides -345 to -330.
AP-1'	Region containing a six out of seven base pair match to an AP-1 consensus sequence. Spans nucleotides -324 to -308.
E-box 1-3	Regions containing E-box consensus sequences. E-box 3 spans nucleotides -308 to -303 (identified by electrophoretic mobility shift assays), E-box 2 spans nucleotides -177 to -155 and E-box 1 spans nucleotides -67 to -47.
Purine-Rich	Region rich in dG and dA nucleotides. Spans nucleotides -284 to -264.
'CRE'	Region displaying homology to both CRE/ATF and AP-1 consensus sequences. Spans nucleotides -198 to -180.
Minimal Element	Region located between the TATA box and the transcriptional start site. Spans nucleotides -20 to -14.
Differential Binding	Region displaying tissue specificity in its protein binding activity. Spans nucleotides +55 to +130.
U.D. 2	Unidentified Element 2. Spans nucleotides +317 to +341.
dT	Region rich in dT nucleotides. The HeLa generated footprint spans nucleotides +372 to +390 and the total rat brain generated footprint spans nucleotides +372 to +405.


dsDBP
ssDBP
AP-1'
AP-1
E-Box 3
E-Box 1
+1
Differential
Binding
-856
+447
Purine-Rich
Minimal
Element
dAT
U.D. 1
dG
'CRE'
E-Box 2
dT
dG
U.D. 2

site. A diagrammatic representation of how these sites are arranged on the promoter is shown in Figure 35.

The AP-1/AP-1'/E-box 3 and 'CRE'/E-box 2 elements appear to represent two domains within the promoter, each spanning approximately 50 bp and separated by approximately 100 bp. Both domains contain E-box motifs and sequences displaying homologies with both AP-1 and/or a CRE/ATF consensus sequences, (AP-1, AP-1' and 'CRE'). The close proximity of the two domains to each other and the possible overlap of proteins binding to sites within the domains raises the possibility that they may synergise with one another to regulate gene expression.

The footprinted region located between these two domains, on the complementary strand, resembles a CT element and finally there is a third E-box motif, termed E-box 1, located approximately 60 bp upstream of the transcriptional start site.

Support for these elements having a functional significance in the regulation of rPPT-A gene expression comes from a comparison of the rat and bovine PPT-A promoter sequences. When the 5' flanking regions of the rat and bovine PPT-A genes are compared a significant degree of homology (82%) is revealed within 346 bp of sequence, immediately upstream of the transcriptional start site, with considerable divergence further upstream (Carter & Krause, 1990; Gilchrist *et al.*, 1991). When the potential regulatory elements identified from the rat promoter, contained within this homologous region, are aligned with the corresponding regions of the bovine promoter a highly significant degree of homology is revealed (Figure 36). The homology between the E-box-1 and E-box-2 motifs, the 'CRE' elements, and the AP-1 and AP-1' motifs is 100%, with the homology between the E-box-3 motifs being 83% and between the CT elements being 77%.

Such a high degree of sequence conservation between 5' flanking regions of the rat and bovine PPT-A promoters suggests that this region of the rPPT-A promoter may play an important role in transcriptional regulatory mechanisms.

3.7.3.2. Flanking regions.

Several related elements have been identified within the rPPT-A promoter which are located in regions placed both 5' and 3' of the major transcriptional start site. These have been termed 5' and 3' G-rich elements, 5' AT- and 3' T-rich elements and Unidentified (U.D.) elements 1 and 2. A diagrammatic representation of how these elements are arranged on the promoter is shown in Figure 37. From this figure it appears that they can be grouped into two domains, one 5' and the other 3' of the major transcriptional start site. Each domain spans approximately 120 bp and contains a G-rich sequence and a T or AT-rich sequence, separated by a U.D. element. The sequence homologies displayed between the 5' and 3' G-rich elements, the 5' AT and 3' T-rich

Figure 35. Arrangement of identified potential regulatory elements within a region of the rPPT-A promoter spanning 320bp 5' of the transcriptional start site.

DNase 1 footprinting analysis was carried out, using the constructs prPPT-βGal6, pVL29 and SM1/Q5 with HeLa nuclear extract, over a region of the rPPT-A promoter spanning nucleotides -40 to -360, relative to the major transcriptional start site. A number of protein bound sequence elements were identified termed AP-1, AP-1', E-box 1 to 3, Purine-Rich and "CRE".

A diagrammatic representation of the rPPT-A promoter, spanning 360 bp of sequence immediately 5' of the major transcriptional start site, is given with identified protein/DNA interaction sites within a region spanning nucleotides -360 to +447 shown. The AP-1/AP-1'/E-box 3 and 'CRE'/E-box 2 elements appear to represent two domains on the promoter, each spanning approximately 50 bp and separated by approximately 100 bp.

AP-1	Region containing a perfect AP-1 consensus sequence. Spans nucleotides -345 to -330.
AP-1'	Region containing a six out of seven base pair match to an AP-1 consensus sequence. Spans nucleotides -324 to -308.
E-box 1-3	Regions containing E-box consensus sequences. E-box 3 spans nucleotides -308 to -303 (identified by electrophoretic mobility shift assays), E-box 2 spans nucleotides -177 to -155 and E-box 1 spans nucleotides -67 to -47.
Purine Rich	Region rich in dG and dA nucleotides. Spans nucleotides -284 to -264.
'CRE'	Region displaying homology to both CRE/ATF and AP-1 consensus sequences. Spans nucleotides -198 to -180.

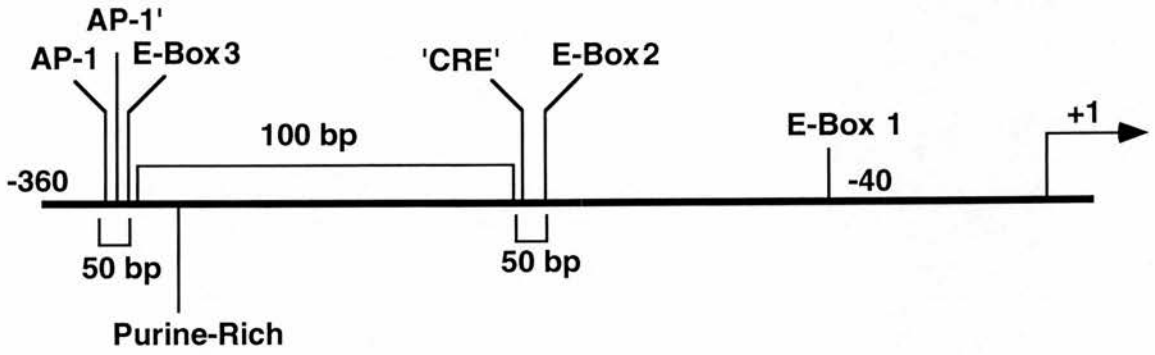

AP-1'
AP-1
E-Box3
'CRE'
E-Box2
E-Box 1
+1
100 bp
-360
-40
50 bp
50 bp
Purine-Rich

Figure 36. A sequence comparison of protein/DNA interaction sites identified within a 5' flanking region of the rat PPT-A promoter with the corresponding regions from the bovine promoter.

A region from the rat PPT-A gene spanning 320bp, 5' of the major transcriptional start site, was analysed for protein/DNA interaction sites by DNase 1 footprinting and electrophoretic mobility shift assays. Several potential regulatory elements were identified which display homologies to the DNA consensus sequences for various transcription factors (shown in bold type).

The identified protein/DNA interaction sites from the rat gene have been compared with the corresponding regions from the bovine PPT-A gene. This comparison reveals a high degree of sequence conservation between the two species.

1. E-box 1.

Rat	-67	AGaGtGT**CACGTG**GcTCtCC	-47
Bovine	-67	AGcGcGT**CACGTG**GgTCcCG	-47

2. E-box 2.

Rat	-177	TTtGGTC**CAGATG**TTATGgaCTC	-155
Bovine	-177	TTgGGTC**CAGATG**TTATGcgCAC	-155

3. E-box 3

Rat	-308	**cAGGTG**	-303
Bovine	-308	**tAGGTG**	-303

4. 'CRE'

Rat	-198	AT**TGCGTCA**TTTCGAACCC	-180
Bovine	-198	AT**TGCGTCA**TTTCGAACCC	-180

5. Purine-rich element

Rat	-284	AAGAAgagGGGagGGGGGCG	-264
Bovine	-284	AAGAAagaGGGgaGGGGGCG	-264

6. AP-1 elements

Rat	-324	tTTTGA**TGAGTAA**TCTC	-308
Bovine	-324	cTTTGA**TGAGTAA**TCTT	-308

Rat	-345	aGcA**TGAGTCA**CTTcG	-330
Bovine	-345	gGaA**TGAGTCA**CTTtG	-330

Figure 37. Arrangement of G-rich, T and AT-rich and unidentified elements 1 and 2 within the rPPT-A promoter.

DNase 1 footprinting analysis of the rPPT-A promoter was carried using the constructs prPPT-βGal4 and pSM1'/Q5 with HeLa, rat brain and rat cerebellar nuclear extracts. A number of sequence elements bound by proteins were identified. These have been termed 5' and 3' G-rich elements, 5' AT-rich and 3' T-rich elements and unidentified elements 1 and 2.

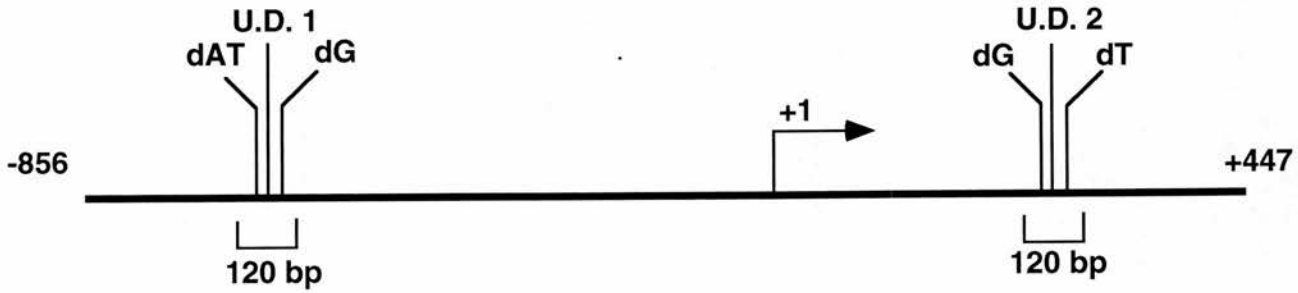
A diagrammatic representation of the rPPT-A promoter, spanning nucleotides -865 to +447, including the sites of protein/DNA interactions is given. These elements appear to represent two domains, one 5' and the other 3' of the major transcriptional start site, each spanning approximately 120 bp of sequence.

dAT Region rich in dA and dT nucleotides which spans nucleotides -662 to -655 and contains an octamer-like consensus sequence (identified by electrophoretic mobility shift assays).

U.D. Unidentified Elements
U.D. 1 spans nucleotides -618 to -607.
U.D. 2 spans nucleotides +317 to +341.

dG Regions rich in dG nucleotides which span nucleotides -575 to -546 and +268 to +290.

dT Region rich in dT nucleotides. The HeLa generated footprint spans nucleotides +372 to +390 and the rat total brain generated footprint spans nucleotides +372 to +405.


U.D. 1
dAT
dG
U.D. 2
dG
dT
+1
-856
+447
120 bp
120 bp

elements and both U.D. elements (Sections 3.7.2. -1.3.b., -1.4. and -1.5.) suggests that similar or related proteins may bind to the 5' and 3' domains and that they may, in some way, synergise with each other to regulate rPPT-A gene expression. Both G-rich elements may potentially be bound by the transcription factor Sp1 (Section 3.7.2.1.3.b.) and as it has been shown that different DNA elements bound by Sp1 molecules may interact by looping out the intervening DNA (Su *et al.*, 1991), it is possible that the rPPT-A 5' and 3' G-rich elements may communicate in such a manner. Additionally, different elements within the 5' and 3' domains may also synergise with one another as, for example, there have been reports in the literature of a functional synergism between the transcription factors Sp1 and Oct-1 involving cooperativity in binding (Janson and Pettersson, 1990). As the AT- and T-rich sites, identified within the rPPT-A promoter, may potentially be bound by a member(s) of the octamer binding protein family, including Oct-1, it is possible that they may interact with the G-rich sites in such a manner.

Section 4: Regulation Of Octamer Binding Proteins By Nerve Growth Factor.

4.1. Introduction.

Octamer binding proteins are a subset of a larger family of homoebox proteins which, in addition to a homeodomain sequence, contain an adjacent conserved sequence, termed a POU-specific domain (Section 1.2.3.1.1.d.). Members of the octamer binding protein family recognise and bind to the octameric DNA consensus sequence ATTTGCAT, first recognised in the histone H2B (Harvey *et al.*, 1982) and heavy and kappa light chain immunoglobulin genes (Falkner and Zachau, 1984). The octamer motif has subsequently been identified as a regulatory sequence motif for a number of ubiquitous and cell specific cellular and viral genes (reviewed in Kemler and Schaffner, 1990). It has been shown that the consensus sequence is highly degenerate (Baumruker *et al.*, 1988) with flanking sequences also being important for mediating protein/DNA interactions.

Numerous POU-domain genes have been identified in mammals and it is likely that many more remain to be discovered. The proteins encoded by the genes isolated so far have been postulated to serve a number of different functions and each show varying and overlapping tissue distributions (He *et al.*, 1989). Each POU-domain protein has been assigned a specific Oct number, although the terminology is very confusing and it appears that similar proteins may be referred to by several different names. The POU-domain proteins discovered to date include the ubiquitously expressed Oct-1 protein (Sturm *et al.*, 1988), the B-cell specific Oct-2 protein (Clerc *et al.*, 1988; Ko *et al.*, 1988; Muller-Immergluck *et al.*, 1988; Scheidereit *et al.*, 1988), Oct-3 which is developmentally regulated during mouse embryogenesis (Okamoto *et al.*, 1990), Oct-4 which is expressed in male and female primordial germ cells and is specific for the female germline at later stages of germ cell development (Scholer *et al.*, 1989, 1990), Oct-5 which is detected in unfertilised oocytes and also in embryonic stem cells (Scholer *et al.*, 1989) and Oct-6 which is expressed in embryonic stem cells, glial progenitor cells and in a restricted set of neurons in the CNS (Suziki *et al.*, 1990).

A number of nervous-tissue specific octamer binding proteins have also been identified, such as Brain-1 (Brn-1), Brain-2 (Brn-2), Brain-3a, 3b and 3c (Brn-3a, 3b and 3c), Brain-4 (Brn-4) and Testes-1 (Tst-1), which show distinct patterns of expression in the developing mammalian central nervous system (He *et al.*, 1989; Hara *et al.*, 1992; Theil *et al.*, 1993; Xiang *et al.*, 1993). Additionally, POU-domain proteins differentially expressed in human glioblastoma and neuroblastoma cell lines have been identified and are termed N-Oct 2α and β, N-Oct 3, N-Oct 4 and N-Oct 5a/5b (Schreiber *et al.*, 1990).

Alternate splicing of primary transcripts generates further heterogeneity among POU-domain proteins. For example, the human Oct-1 primary transcript is alternatively spliced to generate an NH_2- and COOH-terminal truncated form of Oct-1, termed Oct-1b (Das and Herr, 1993). In transient transfection assays Oct-1b displays an enhanced ability to activate the human histone H2B promoter than the larger Oct-1 protein (Das and Herr, 1993). There are also alternative Oct-2 transcripts, termed Oct-2a, Oct-2b (Schreiber *et al.*, 1988), Oct 2c and Oct-2d (Mini-Oct) (Stoykova *et al.*, 1992). POU-domain proteins are also thought to be both positively and negatively regulated through protein-protein interactions (Section 1.2.3.1.1.d.).

Lindsay and Harmar (1989) have shown that the levels of rPPT-A transcripts in adult DRG neurons in culture increase with nerve growth factor (NGF), a trophic molecule essential for the survival of sympathetic and sensory neurons. Subsequently, by *in situ* hybridisation, Henken *et al.* (1990) showed that levels of rPPT-A transcripts in adult DRG neurons may depend on the availability of NGF. Therefore, it seems possible that some of the transcription factors regulating expression of the rPPT-A gene may also be regulated by NGF. One such candidate is the AP-1 complex which has already been shown to vary in composition in PC12 cells with NGF treatment (Curran and Morgan, 1985; Greenberg *et al.*, 1985; Kruijer *et al.*, 1985; Milbrandt, 1986; Quinn, 1991). Another potential candidate for an NGF regulated transcription factor is a member of the octamer binding protein family.

Evidence exists which suggests that there may be a link between octamer binding proteins and NGF in the regulation of HSV immediate early gene expression. The first stage of the lytic cycle following HSV infection involves the transcription of the immediate early (IE) genes, ICP0, ICP4, ICP22, ICP27 and ICP47, whose products are essential for subsequent stages of viral gene expression (Preston, 1979; Watson & Clements, 1980). HSV can establish latent infection in sensory neurons (reviewed by Roizman & Sears, 1986) and *in situ* hybridisation studies have reported that none of the five IE gene mRNAs are detected in latently infected ganglia (Stevens *et al.*, 1987). The maintenance of HSV latency in *in vitro* systems has been shown to require NGF. If anti-NGF antibodies are added to primary cultures of rat sympathetic neurons, the virus reactivates and cells begin producing viral antigens (Wilcox and Johnson, 1987, 1988).

Transcription of immediate early genes involves the interaction of various cellular transcription factors, notably Sp1 and Oct-1, with sites on IE gene promoters (Jones and Tjian, 1985; O'Hare and Goding, 1988). Oct-1 interacts, in combination with the HSV encoded protein VP16, with the octamer-like TAATGARAT motif on immediate early promoters (O'Hare & Goding, 1988). The failure of immediate early gene expression in latently infected neuronal cells has been proposed to be due to a

difference in the nature or type of transcription factors in such cells. C1300 mouse neuroblastoma cells have been used as a model system for studying the interaction of HSV with neuronal derived cells (Kemp and Latchman, 1989; Kemp *et al.* 1990) and it has been demonstrated that there is a difference in the nature of octamer binding proteins in such cells. C1300 cells were found to express an additional octamer binding protein, absent from non-neuronal cells, and this protein was proposed to mediate the repression of IE promoters in neuronal cells (Kemp *et al.*, 1990). As NGF is required to maintain latency, another mechanism for altering transcription from IE promoters may be NGF modulation of some of the transcription factors involved in IE gene transcription. A potential candidate for such modulation may be a member of the octamer binding protein family.

If NGF does indeed regulate the nature of octamer binding proteins then this phenomena may be important for mediating the effects of NGF on rPPT-A gene expression through octamer binding elements located within the promoter.

4.2. Aim of the study.

The aim of this study was to investigate the potential regulation by NGF of octamer binding proteins which in turn may regulate the rPPT-A promoter.

4.3. Strategy.

In order to investigate the possible regulation of octamer binding proteins by NGF, electrophoretic mobility shift assays were carried using out as described in Sections 2.2.6.1. and 3.6. Unfortunately, it is difficult to generate sufficient protein extract of suitable quality from primary cultures of DRG neurons for this type of study to be carried out. Therefore, this study has made use of the C1300 mouse neuroblastoma cell line (Augusti-Tocco and Sato, 1969) which has been previously been used in other studies as a model system for cells of neuronal origin (Kemp *et al.*, 1990).

4.4. Results.

The DNA consensus sequence for the octamer binding protein from the immunoglobulin heavy-chain enhancer, (Augereau and Chambon, 1986; Staudt *et al.*, 1988) (Figure 38, Oct), was used to identify specific octamer binding protein complexes in C1300 cells by electrophoretic mobility shift assay.

Figure 39, lanes 1-5 demonstrate the presence of three complexes formed between C1300 cells and the Oct oligonucleotide. The two lower complexes were competed by increasing concentrations of a non-specific oligonucleotide, (NS, Figure 39) (lanes 2-5), so classing them as non-specific binding complexes (indicated as

Oct 5' GAGATCTAGCATGCAAATCATTGT
TAGATCGTACGTTTAGTAACAGAG

NS 5' ATCCCTTTAAATTTGCGAGCT
GGAAATTTAAACGCTCGACTA

Figure 38. Sequence of the oligonucleotides used in electrophoretic mobility shift assays.

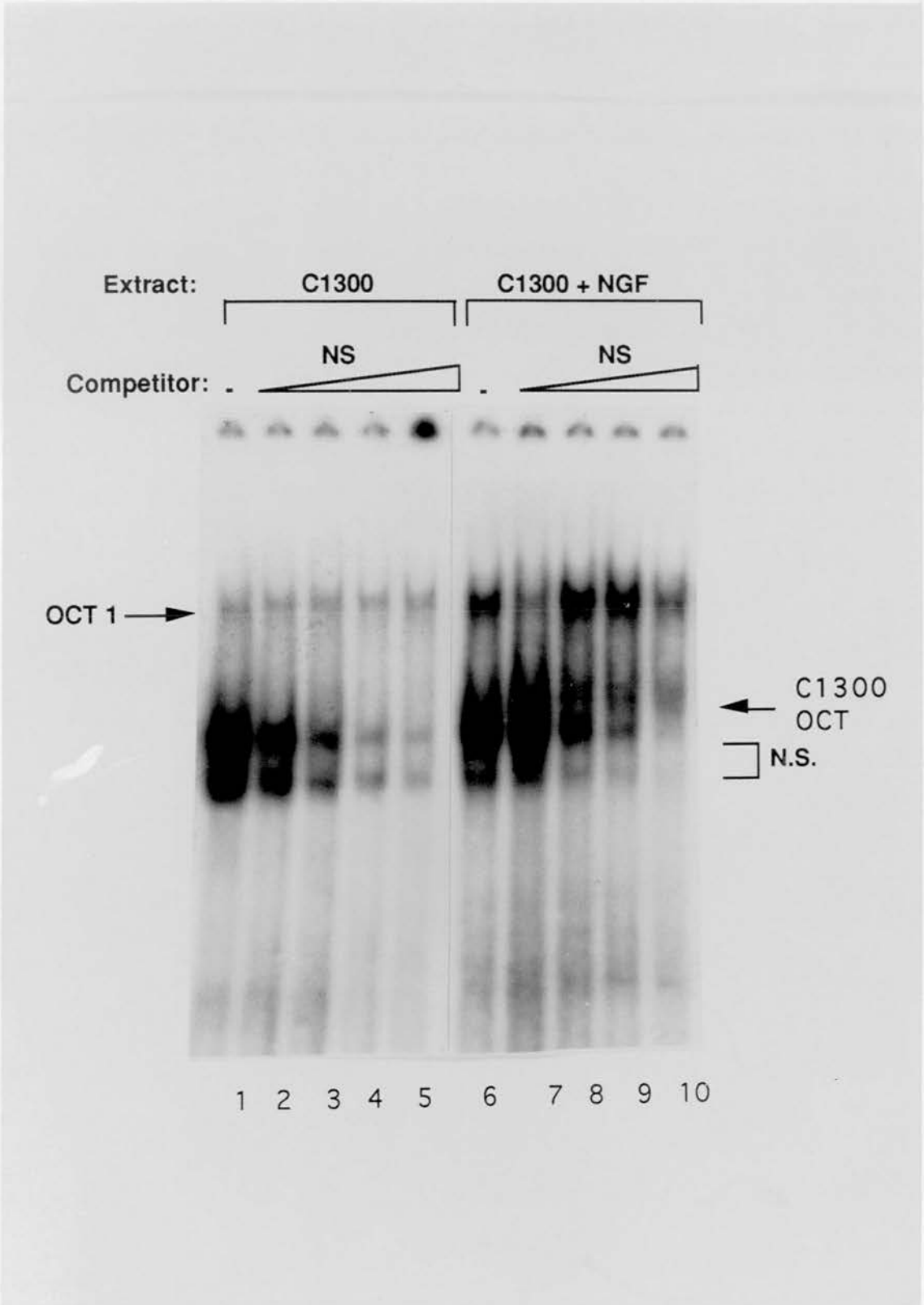
The sequence of the oligonucleotides, Oct (or Specific, S) and NS, used in electrophoretic mobility shift assays are shown.
Oct contains an octamer binding protein element from the immunoglobulin heavy-chain enhancer (Augereau and Chambon, 1986; and Staudt *et al.*, 1988).
NS (Non-specific) is a randomly computer generated sequence.

Figure 39. Electrophoretic mobility shift assay using the immunoglobulin octamer binding element and extract from C1300 cells and C1300 cells exposed to NGF for 24 h.

A double stranded oligonucleotide containing the immunoglobulin octamer binding element (Oct) was 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotide (1ng) was incubated with either extract from C1300 cells (20μg) or with extract from C1300 cells exposed to NGF for 24 h (20μg), in the presence of 1μg of poly d(I-C).

Several complexes were formed between the probe and C1300 cell extract, in the absence of competitor DNA (lane 1). The two lower complexes, indicated as NS, were shown to represent non-specific binding complexes as they were competed by increasing concentrations (10, 50, 100 and 150ng) of non-specific oligonucleotide (NS) (lanes 2-5). The upper complex migrates to a similar position as the Oct-1 complex from HeLa cell extract (see Figure 42) and so is indicated by an arrow to represent Oct-1.

Extract from C1300 cells exposed to NGF for 24 h (lanes 6-10) generated, in addition to the Oct-1 complex, a complex termed C1300-Oct (indicated by an arrow). Increasing concentrations (10, 50, 100 and 250ng) of non-specific oligonucleotide, (NS) (lanes 7-10), competed away only the non-specific binding complexes, thus allowing the C1300-Oct complex to be observed more clearly.


Extract:
C1300
C1300 + NGF
Competitor:
-
NS
-
NS
OCT 1
C1300
OCT
N.S.
1 2 3 4 5 6 7 8 9 10

N.S.). The slower migrating complex was not competed by the NS oligonucleotide and is indicated by an arrow to be the ubiquitous octamer binding protein, Oct-1, as its migration is in the same position as the Oct-1 complex from HeLa cells (Figure 41, lane 7). Lanes 6-10 show that on exposure of C1300 cells to NGF for 24 h a new complex, C1300-Oct (indicated by an arrow), was observed. The C1300-Oct complex was observed more clearly when increasing concentrations of the NS oligonucleotide were added (lanes 7-10).

The specificity of the induced C1300-Oct complex was demonstrated by competition with homologous (S) and heterologous (NS) oligonucleotides (Figures 40 and 41). In these and following assays, cell extracts were preincubated with 1μg of poly[d(I-C)] and non-specific oligonucleotide (50ng) in order to remove non-specific complexes and to allow the induced complex to be seen more clearly. Increasing concentrations of homologous competitor (S) were added to extract from C1300 cells exposed to NGF for 24 h, (Figure 40, lanes 2-5 and Figure 41, lanes 5-6). This resulted in competition of both the Oct-1 complex and the C1300-Oct complex (indicated by arrows). Increasing concentrations of heterologous competitor, (NS) (Figure 40, lanes 7-10, and Figure 41, lanes 2-3), failed to compete for formation of either the Oct-1 complex or the C1300-Oct complex. The complex generated by ubiquitous Oct-1 from HeLa nuclear extract is shown in Figure 41, lane 7 to demonstrate that it migrates to the same position as the complex generated by Oct-1 from C1300 whole cell extract.

In order to determine the time course over which the C1300-Oct complex was induced, an electrophoretic mobility shift assay was carried out using extracts from C1300 cells which had been treated with NGF for increasing amounts of time. Figure 42, lane 1 shows the ubiquitous Oct-1 complex (indicated by an arrow) generated by HeLa cell extract. Lane 2 contains C1300 cell extract and lanes 3-5 contain extracts from C1300 cells treated with NGF for 1, 3 and 24h respectively. As expected, Oct-1 was constitutively present in all of the C1300 cell extracts. The C1300-Oct complex (indicated by an arrow) appeared as a faint complex after exposure to NGF for 3 h (lane 4) and was clearly observed at 24 h post-induction (lane 5).

Previous reports have indicated that C1300 cells constitutively express two specific octamer binding proteins (Kemp *et al.*, 1990). In order to demonstrate that the finding of an NGF inducible protein in C1300 cells was not unique to one particular isolate of cells, a different source of C1300 cells, from a different clone, was obtained. Figure 43 shows that the results are reproducible. Oct-1 (indicated by an arrow) was constitutively present in C1300 cells (lanes 1, 2, 5, 6 and 9-11). The C1300-Oct complex (indicated as an arrow) appeared after exposure to NGF, in this case for 48 and 72 h (lanes 5, 6 and 9-11). The increase in abundance of the C1300-Oct complex at

Figure 40. Electrophoretic mobility shift assay using the immunoglobulin octamer binding element and extract from C1300 cells exposed to NGF for 24 h.

A double stranded oligonucleotide containing the immunoglobulin octamer binding element (Oct) was 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotide (1ng) was incubated with extract from C1300 cells exposed to NGF for 24 h (20μg), in the presence of 1μg of poly d(I-C) and 50ng of non-specific oligonucleotide.

Several complexes were formed between the probe and C1300 cell extract, in the absence of competitor DNA (lanes 1 and 6). Two complexes, indicated as Oct-1 and C1300-Oct, were shown to be specific as they were competed by increasing concentrations (10, 25, 50 and 100ng) of homologous oligonucleotide competitor (S) (lanes 2-5). Increasing concentrations (10, 50, 100 and 150ng) of a heterologous oligonucleotide, (N.S) (lanes 7-10), competed only the lower non-specific complexes (indicated as NS).

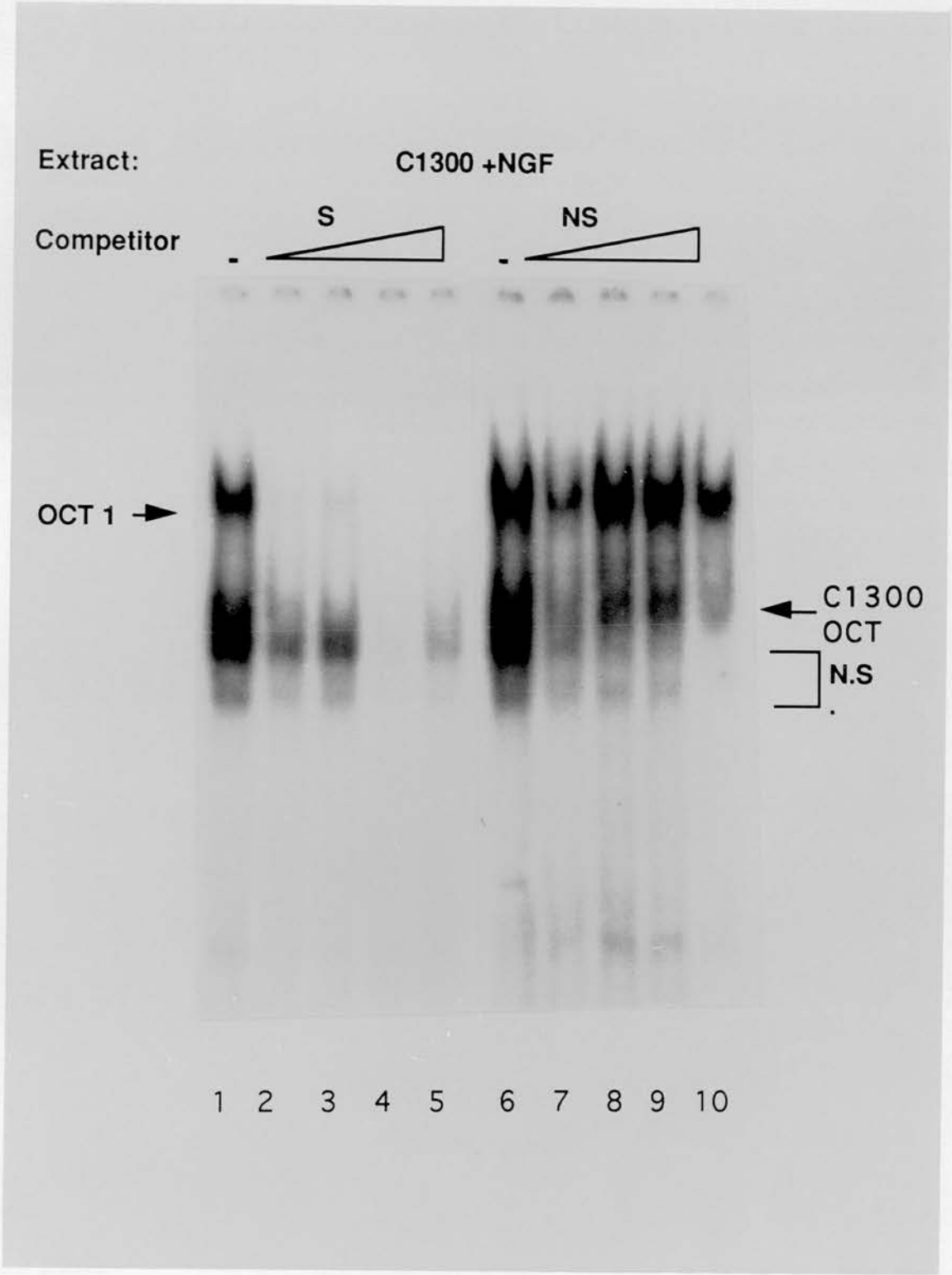

Extract:
C1300 +NGF
Competitor
S
NS
-
-
OCT 1
C1300
OCT
N.S
1 2 3 4 5 6 7 8 9 10

Figure 41. Electrophoretic mobility shift assay using the immunoglobulin octamer binding element and extract from HeLa cells and C1300 cells exposed to NGF for 24 h.

A double stranded oligonucleotide containing the immunoglobulin octamer binding element (Oct) was 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotide (1ng) was incubated with either HeLa nuclear extract or extract from C1300 cells exposed to NGF for 24 h (20μg), in the presence of 1μg of poly d(I-C) and 50ng of non-specific oligonucleotide.

In lane 7 the Oct-1 complex formed between the probe and HeLa nuclear extract is indicated by an arrow. Lanes 1-6 contain C1300 extract treated with NGF for 24 h. Several complexes were formed between the probe and C1300 cell extract, in the absence of competitor DNA (lanes 1 and 4). Two complexes, indicated as Oct-1 and C1300-Oct, were shown to be specific as they were not competed by increasing concentrations (25 and 50ng) of a heterologous oligonucleotide, (NS) (lanes 2 and 3), but they were competed by increasing concentrations (25 and 50ng) of homologous oligonucleotide (S) (lanes 5 and 6).

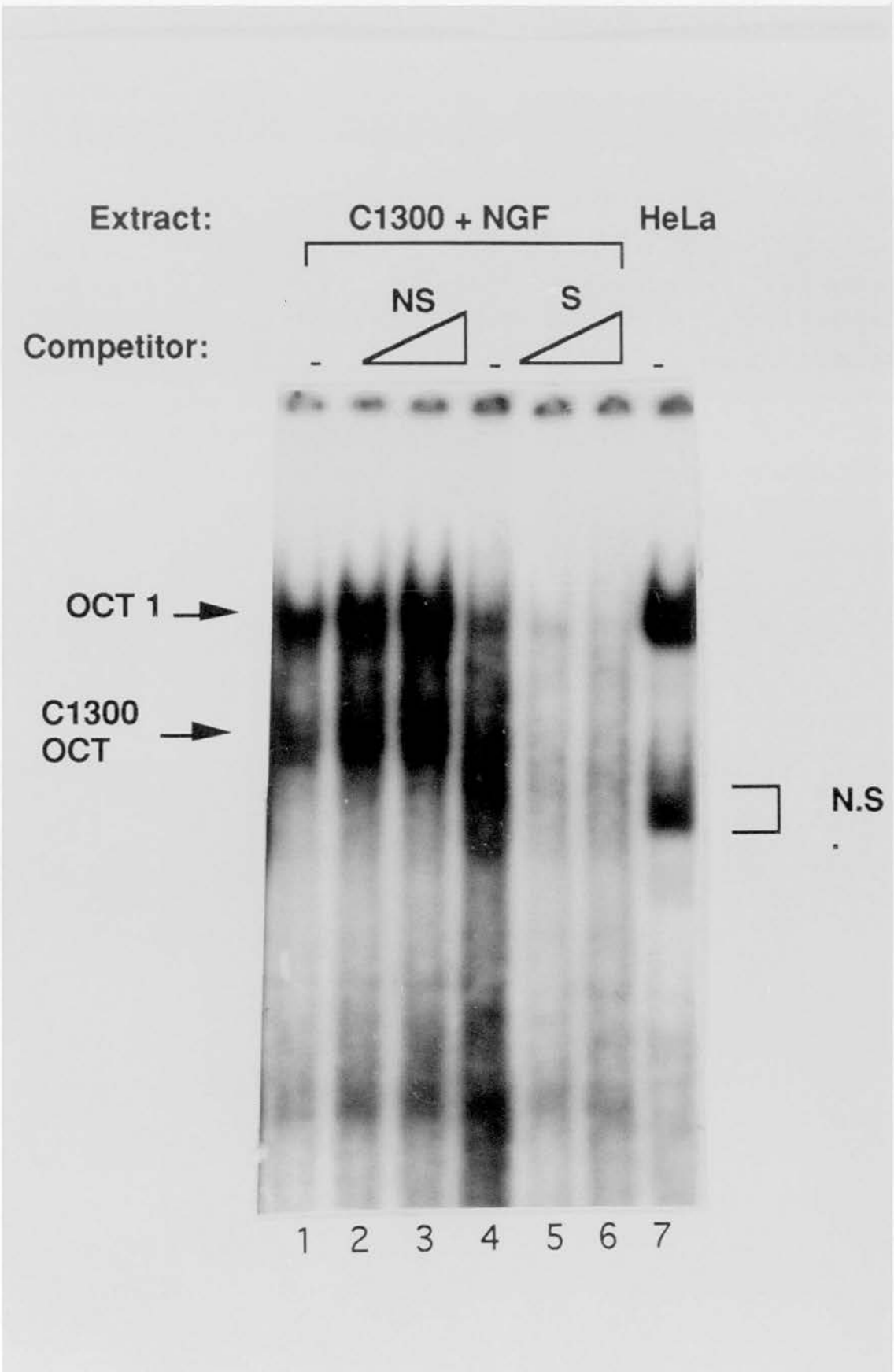

Extract:
C1300 + NGF
HeLa
NS
S
Competitor:
-
-
-
OCT 1
C1300
OCT
N.S
1 2 3 4 5 6 7

Figure 42. Electrophoretic mobility shift assay using the immunoglobulin octamer binding element and extract from HeLa cells and C1300 cells exposed to NGF for 1, 3 or 24 h.

A double stranded oligonucleotide containing the immunoglobulin octamer binding element (Oct) was 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotide (1ng) was incubated with either HeLa nuclear extract (20μg) or with extract from C1300 cells exposed to NGF for 1, 3 and 24 h (20μg), in the presence of 1μg of poly d(I-C) and 50ng of non-specific oligonucleotide.

Lane 1 shows the Oct-1 complex (indicated by an arrow) formed between the probe and HeLa nuclear extract, in the absence of competitor DNA. Lane 2 contains extract from C1300 cells and lanes 3-5 contain extract from C1300 cells treated with NGF for 1, 3 and 24 h respectively. The C1300-Oct complex (indicated by an arrow) appeared as a faint complex after 3 h (lane 4) and was clearly visible after 24 h of NGF treatment (lane 5).

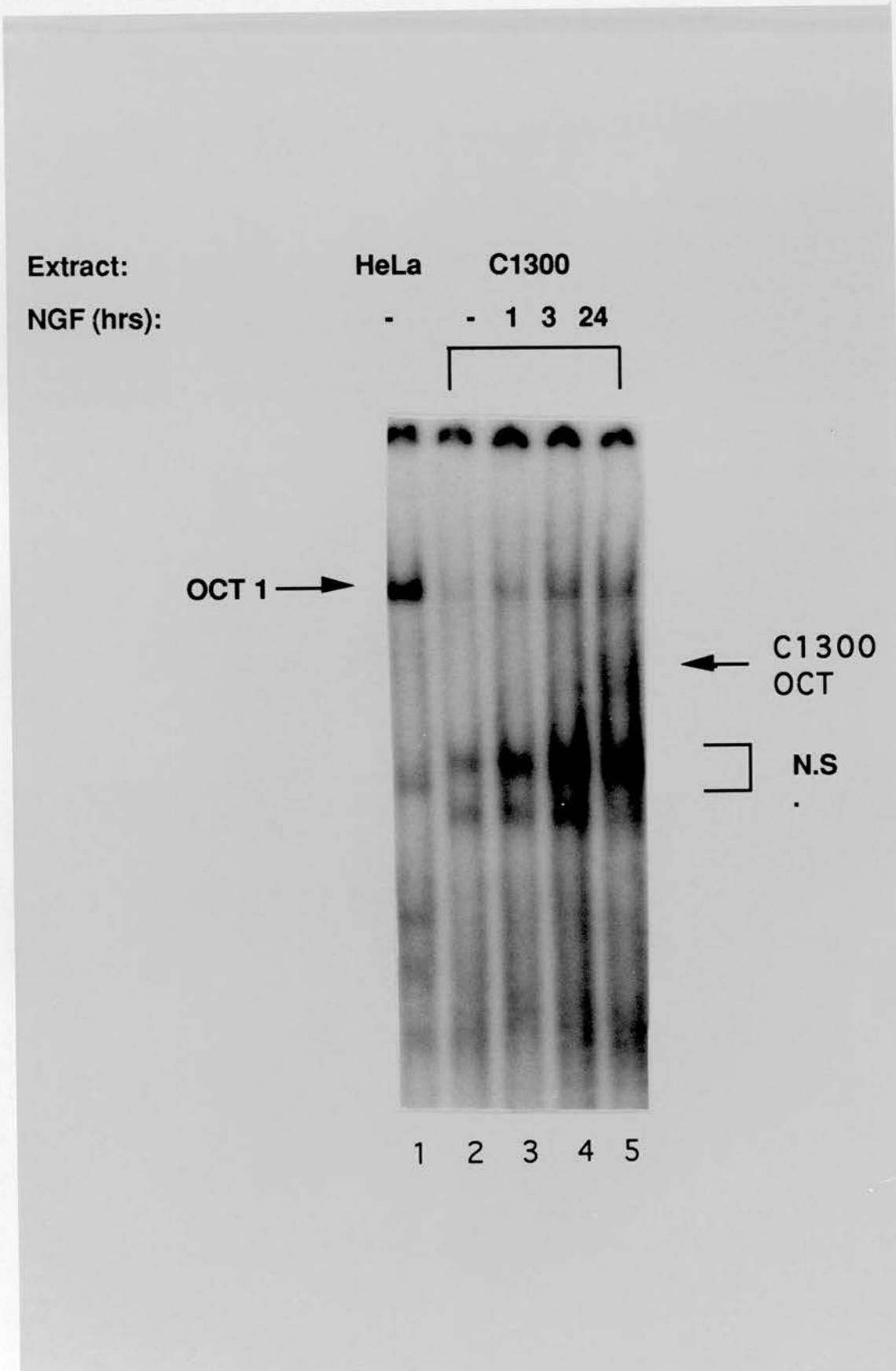
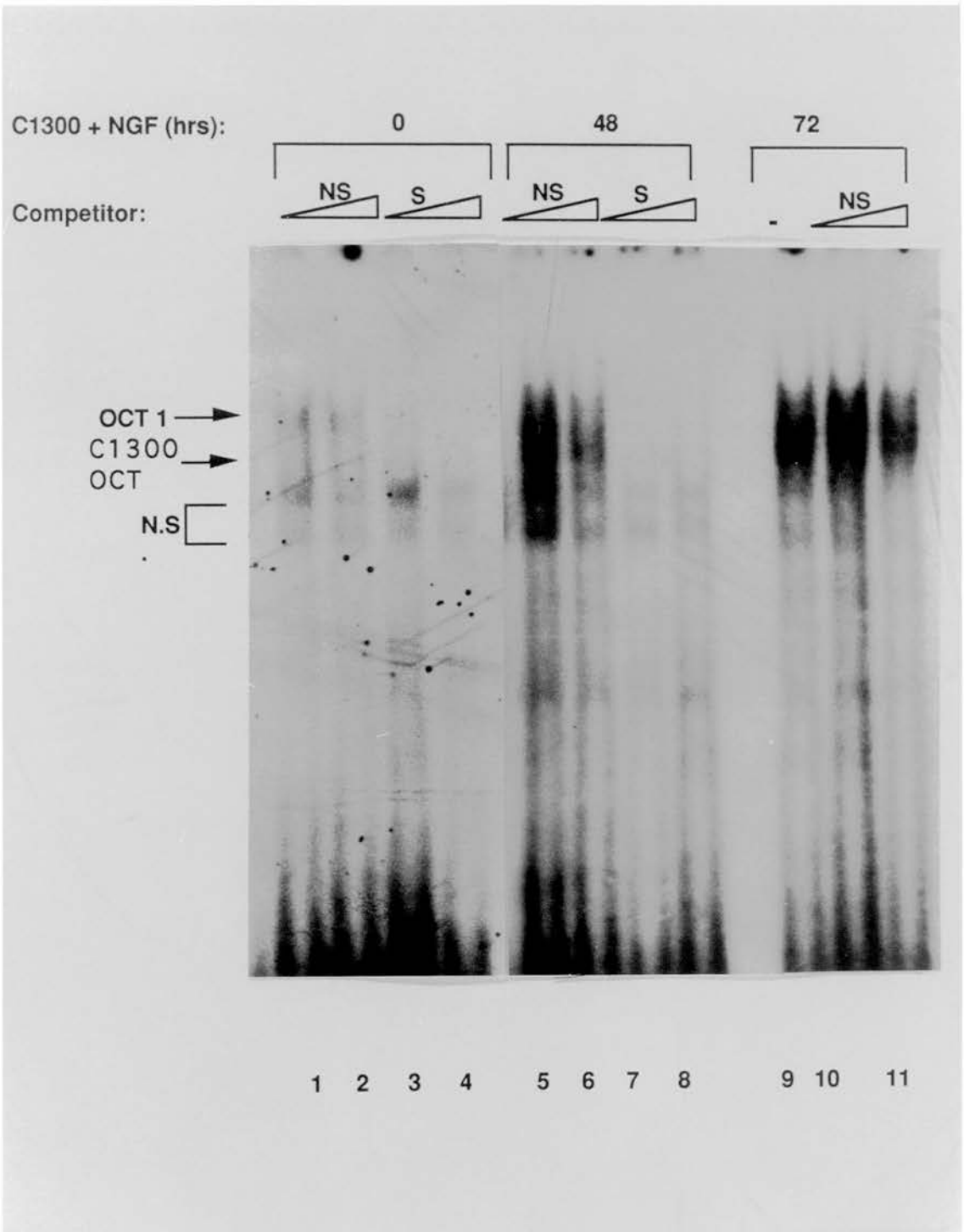

Extract:
HeLa
C1300
NGF (hrs):
-
-
1
3
24
OCT 1
C1300
OCT
N.S
.
1
2
3
4
5

Figure 43. Electrophoretic mobility shift assay using the immunoglobulin octamer binding element and extract from C1300 cells and C1300 cells exposed to NGF for 48 or 72 h.

A double stranded oligonucleotide containing the immunoglobulin octamer binding element (Oct) was 5' end labelled with T4 polynucleotide kinase. End-labelled oligonucleotide (1ng) was incubated with either extract from C1300 cells (20μg) or with extract from C1300 cells (20μg) exposed to NGF for 48 and 72 h (provided by J.Morrow), in the presence of 1μg of poly d(I-C) and 50ng of non-specific oligonucleotide.

In lanes 1-4 the probe was incubated with extract from C1300 cells, in lanes 5-8 with extract from C1300 cells exposed to NGF for 48 h and in lanes 9-11 with extract from C1300 cells exposed to NGF for 72 h. A complex corresponding to Oct-1 (indicated as an arrow) was formed by all cell extracts which was competed by 100ng of a homologous oligonucleotide, (S) (lanes 3, 4, 7 and 8), but not by 100ng of a heterologous competitor (NS) (lanes 1, 2, 5, 6, 10 and 11). Only cells treated with NGF formed the C1300-Oct complex (indicated as an arrow). The C1300-Oct complex was competed by 100ng of a homologous oligonucleotide, (S) (lanes 7 and 8), but not by 100ng of a heterologous competitor (NS) (lanes 5, 6, 10 and 11).


C1300 + NGF (hrs):
0
48
72
Competitor:
NS
S
NS
S
-
NS
OCT 1
C1300
OCT
N.S
1 2 3 4 5 6 7 8 9 10 11

48 and 72 h post-induction was such that it obscured the Oct-1 complex. The specificity of the complexes formed was again demonstrated by competition of Oct-1 and C1300-Oct with homologous (S) competitor (lanes 3, 4, 7 and 8) but not by heterologous (NS) competitor (lanes 1, 2, 5, 6, 10 and 11).

4.5. Discussion.

The results from this study demonstrate, by electrophoretic mobility shift assays, that on exposure of C1300 mouse neuroblastoma cells to NGF a DNA-binding protein recognising the octamer binding protein consensus sequence is induced.

The protein which appears after exposure of C1300 cells to NGF is termed C1300-Oct and was shown to be absent, or only minimally present, prior to induction (Figure 39). The C1300-Oct protein was shown to appear after treatment of the cells with NGF for 3 h (Figure 42) and progressively increases in abundance with exposure for 24, 48, and 72 h (Figure 42 and Figure 43).

Thus, this study demonstrates that C1300 cells constitutively express only one octamer binding protein, identified as Oct-1 from its migration in electrophoretic mobility shift assays to the same position as Oct-1 from HeLa cells (Figure 41). These findings are in contrast to those of Kemp *et al* (1990) where two distinct octamer binding proteins were shown to be constitutively present in C1300 cells. In order to rule out the possibility that the findings from this study were unique to one particular isolate of cells, a different source were used in a similar assay and a consistent result was produced (Figure 43).

It has been shown that C1300 cells display specific binding of NGF to their surface membranes (Revoltella *et al.*, 1974; Greene and Shooter, 1980) and changes in the differentiation state of C1300 cells, as determined by histochemical staining for acetylcholinesterase expression, have been reported in response to NGF treatment (Brodeur and Goldstein, 1976). Therefore, a change in the expression of particular transcription factors, such as octamer binding proteins, may be expected in C1300 cells in response to NGF treatment.

The identity of the NGF-inducible octamer binding protein, identified in this study, is presently unknown. It is distinct from the ubiquitous Oct-1 protein, as demonstrated by their different migrations in electrophoretic mobility shift assays, however it does display a similar migration to that of purified Oct-2 protein (personal communication, J. Quinn). Interestingly, the neuronal specific octamer binding proteins N-Oct 2α and β (Schreiber *et al.*, 1990) also display similar mobilities to C1300-Oct, suggesting that C1300-Oct may be similar or related to these proteins. Support for this suggestion comes from a study by Wood *et al.* (1992), showing that NGF upregulates the expression of a neuronal form of Oct-2 in DRG neurons.

Determination of the exact nature and expression pattern of the C1300-Oct protein will require further studies to be carried out using specific anti-sera directed against various members of the octamer binding protein family, including Oct-2, in order to determine if any will cross-react with C1300-Oct.

Regulation by NGF has been reported previously for a number of transcription factors. These include NGFI-A and NGFI-B which were both originally isolated based on their rapid induction by nerve growth factor in the rat pheochromocytoma cell line, PC12 (Milbrandt, 1987, 1988). The products of the proto-oncogenes, c-fos and c-jun, have also been shown to be induced rapidly and transiently in response to NGF in PC12 cells (Curran and Morgan, 1985; Greenberg *et al.*, 1985; Kruijer *et al.*, 1985; Milbrandt, 1986). The genes encoding these factors are implicated in mediating the 'early response' to NGF, they are induced to maximal levels within 30 to 60 mins and return to basal levels within about 4 h, and thus are termed immediate early genes. In contrast to these immediate early gene products, C1300-Oct is not detected by electrophoretic mobility shift assay until 3 h post-induction and is present long after the initial stimulus. Therefore, whereas the immediate early gene products are proposed to initiate an early cascade of gene activation in mediating the response to NGF, the C1300-Oct protein may act to mediate some of the long term responses to NGF.

It has been shown that the immediate early gene induction by NGF is not specific to NGF. Epidermal growth factor (EGF) and NGF have both been shown to induce the same set of immediate early genes in PC12 cells (Greenberg *et al.*, 1985; Bartel *et al.*, 1989), yet they promote distinct biological effects. EGF promoted effects are primarily proliferative, while NGF is a differentiative factor. Selectivity in the action of NGF has been proposed to occur at several steps in the signalling cascade, including at the receptor level and at later gene activities (reviewed by Chao, 1992). A number of specific genes primarily encoding structural proteins and enzymes which reflect a neuronal phenotype have been found to be induced in the long term by NGF (reviewed by Chao, 1992) and may provide some selectivity for NGF action. It is possible, therefore, that C1300-Oct is responsible for mediating some of these later steps in the response to NGF by regulating transcription from such neuronal phenotype-determining genes.

Interestingly, two potential octamer binding elements have been identified within the rPPT-A promoter (Sections 3.7.2.1.4.). The identity of the factor(s) which will bind to these elements is unknown but, as the levels of rPPT-A transcripts in adult sensory neurons have been shown to depend on NGF availability (Lindsay and Harmar, 1989; Henken *et al.*, 1990), it is possible that NGF-inducible octamer binding proteins, including C1300-Oct, may be involved in the regulation of rPPT-A gene expression. Electrophoretic mobility shift assays, using the two octamer-like elements

from the rPPT-A promoter as oligonucleotide probes and extract from C1300 cells and C1300 cells treated with NGF, should be carried out in order to determine if NGF treatment alters the pattern of proteins which will bind to the two elements.

Section 6: Summary and Future Directions.

6.1. Summary.

The work presented in this thesis describes the identification of multiple potential regulatory elements within a region of the rPPT-A promoter, spanning nucleotides -865 to +447, previously shown to be sufficient for expression of a reporter gene in microinjected DRG neurons (Mulderry *et al.*, 1993).

By the identification of these elements, this study has implicated the involvement of several families of transcription factors in the regulation of rPPT-A gene expression. The majority of the elements were identified using HeLa cell extract and it is expected that DRG neurons, where the rPPT-A gene is expressed endogenously at high levels, will express a distinct array of transcription factors belonging to these different families of DNA-binding proteins. Therefore, while some of the identified elements may be recognised by proteins present in DRG neurons, it is possible that such tissue will also express proteins which recognise a distinct set of elements from those recognised by HeLa cell proteins. Such differential binding by proteins from different cell or tissue sources to regulatory elements within the rPPT-A promoter may potentially allow for tissue-specific gene regulation. Evidence to support such tissue-specific binding comes from the finding, in this study, that some of the elements within the rPPT-A promoter, located 3' of the transcriptional start site, differ in their binding characteristics depending on whether extracts of neuronal or non-neuronal tissue are used. Other elements which may play a role in mediating tissue-specific expression of the rPPT-A gene include a CCAAT element, spanning nucleotides -181 to -177, previously noted by sequence analysis (Carter and Krause, 1990). DNase 1 footprinting analysis, utilised in this study, was unable to detect binding over this sequence element by proteins expressed in HeLa cells. CCAAT elements are present in a large number of cellular promoters (Christy *et al.*, 1989; Friedman *et al.*, 1989; Crossley and Brownlee, 1990) and are bound by CAAT box/enhancer binding proteins (C/EBPs). These proteins belong to a large family of bZIP proteins and have been shown to display a wide and varied pattern of expression (Landschulz *et al.*, 1989; Cao *et al.*, 1991; Williams *et al.*, 1991). It is possible, therefore, that DRG neurons express a distinct set of C/EBPs from those found in HeLa cells and that these will display a different pattern of binding to regulatory elements within the rPPT-A promoter, thus potentially allowing for tissue-specific gene expression.

In addition to an involvement in tissue-specific gene regulation, the identified elements are also likely to have an important role to play in the regulation of gene expression in response to various extracellular stimuli. A number of factors have been identified as potential regulators of the rPPT-A gene (Section 1.1.5.). One such factor

is NGF which has been shown to increase the levels of rPPT-A mRNA in adult rat DRG neurons (Lindsay and Harmar, 1989). It is possible that this increase in mRNA levels is mediated through one or a number of the elements identified in this study. Potential candidates include the identified AP-1 or AP-1 like elements, AP-1, AP-1', HeLa AP-1 and 'CRE', as it has been shown that the levels of proteins which bind to AP-1 elements increase or vary in response to NGF treatment (Curran and Morgan, 1985; Greenberg *et al.*, 1985; Kruijer *et al.*, 1985; Milbrandt, 1986; Quinn, 1991). Other potential mediators of NGF actions include the identifed E-box motifs as it has been shown the levels of the E-box binding protein MASH-1 are increased in PC12 cells with NGF treatment (Johnson *et al.*, 1990, 1992). The 5' AT-rich and 3' T-rich octamer-like sites may also be important for mediating the transcriptional responses to NGF , as it has been demonstrated that the levels of N-Oct 2 increase in DRG neurons in response to NGF (Wood *et al.*, 1992). Additionally, it has been shown in this study that, in C1300 neuroblastoma cells, a DNA binding protein recognising the octamer binding element from the immunoglobulin enhancer is induced by NGF.

Therefore the multiple potential regulatory elements identified in this study, within a 1300 bp region of the rPPT-A promoter, are likely to be important for mediating many of the transcriptional responses of the rPPT-A gene.

6.2. Future directions.

The results obtained from this study, describing the identification of multiple potential regulatory elements within the rPPT-A promoter, will now lead on to a number of future studies aimed at furthering our understanding of the regulation of the rPPT-A gene at the transcriptional level.

Since the work presented in this thesis was carried out, there have been reports of a rat pancreatic endocrine cell line RINm5F which endogenously expresses the PPT-A gene (McGregor *et al.*, 1992). The elements identified in this study can now be further analysed using cells which express SP. Such studies may include further DNase 1 footprinting and electrophoretic mobility shift assays. Additionally, electrophoretic mobility shift assays using fractionated protein extracts from RINm5F cells should allow for the purification and identification of factors which interact with the rPPT-A promoter.

The results presented in this thesis may also be used as the basis for a number functional studies designed to investigate the role(s) of identified elements in regulating rPPT-A gene expression. Such studies may include transfection or microinjection of different rPPT-A gene reporter gene constructs into clonal cell lines (RINm5F) or DRG neurons in culture. By the use of constructs containing individual elements or groups of elements, these studies should determine which elements can function independently or

those which function as part of a group. Exposure of transfected cells or neurons to various extracellular stimuli, such as NGF, should also help to determine which elements are important for conferring the transcriptional response to such stimuli. Site specific mutagenesis of the elements may also be carried out in order to determine their individual importance in regulating rPPT-A gene expression.

Future studies, arising from the results presented in this thesis, may also include the isolation of regulatory proteins present in DRG neurons which will bind to the identified elements. In order to achieve this a DRG cDNA library may be screened using the elements as oligonucleotide probes. Once candidate clones have been identified their roles in the regulation of rPPT-A gene expression in DRG neurons may be investigated. For example, the role of AP-2, which may bind to the rPPT-A promoter (Section 3.7.2.1.3.), in regulating both endogenous rPPT-A gene expression and that of reporter constructs, containing various regions from the promoter, may be investigated. This may be achieved by microinjecting DRG neurons in culture with constructs encoding either wild type AP-2 or an alternative AP-2 transcript, termed AP-2B, which lacks the dimerization domain necessary for DNA-binding and therefore inhibits the binding of endogenous AP-2 to its consensus sequence (Buettner *et al.* 1993). In a similar manner, expression vectors for other candidate factors may also be microinjected into DRG neurons and the importance of each in regulating gene expression may be determined by assaying the expression of either the endogenous rPPT-A gene or of reporter gene constructs. The possible regulation of candidate factors by NGF may also be investigated by assaying gene expression following NGF treatment.

By the identification of potential regulatory elements within the rPPT-A promoter, it is hoped that novel targets for the development of drugs that modulate rPPT-A gene expression may be identified. Substance P is thought to play a role in many processes of clinical relevance, including the mediation of pain perception and in neurogenic inflammation in conditions such as rheumatoid arthritis. Therefore, it is hoped that, by furthering our understanding of the mechanisms involved in regulating rPPT-A gene expression, new approaches may be developed for the treatment of relevant disorders including those involving pain and inflammation.

References.

Adams, C.C., and J.L. Workman (1993). Nucleosome displacement in transcription. Cell, 72, 305-308.

Adhya, S. (1989). Multipartite genetic control elements: communication by DNA loop. Annual Review of Genetics, 23, 227-250.

Affolter, M., A. Percival-Smith, M. Muller, W. Leupin, and W.J. Gehring (1990). DNA binding properties of the purified Antennapedia homeodomain. Proceedings of the National Academy of Sciences. USA., 87, 4093-4097.

Agre, P., P.F. Johnson, and S. McKnight (1989). Cognate DNA binding specificty retained after leucine zipper exchange between GCN4 and C/EBP. Science, 246, 922-925.

Akagi, H., M. Otsuka, and M. Yanagisawa (1980). Identification by high-performance liquid chromotography of immunoreactive substance P released from isolated rat spinal cord. Neuroscience Letters, 20, 259-263.

Allison, L.A., M. Moyle, M. Shales, and C.J. Ingles (1985). Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell, 42, 599-610.

Allison, L.A., J.K.C. Wong, D. Fitzpatrick, M. Moyle, and C.J. Ingles (1988). The C-terminal domain of the largest subunit of RNA polymerase of *Sacharomyces cerevisiae*, *Drosophila melanogaster* and mammals: a conserved structure with an essential function. Molecular and Cellular Biology, 8, 321-329.

Amati, B., S. Dalton, M.W. Brooks, T.D. Littlewood, G.I. Evans, and H. Land (1992). Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. Nature, 359, 423-426.

Amati, B., M.W. Brooks, N. Levy, T.D. Littlewood, G.I. Evan, and H. Land (1993). Oncogenic activity of the c-Myc protein requires dimerization with Max. Cell, 72, 233-245.

Anderson, W.F., D.H. Ohlendorf, Y. Takeda, and B.W. Matthews (1981). Structure of the cro repressor from λ bacteriophage and its interaction with DNA. Nature, 290, 754-758.

Andrews, N.C., H. Erdjument-Bromage, M.B. Davidson, P. Tempst, and S.H. Orkin (1993). Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. Nature, 362, 722-728.

Angel, P., M. Imagawa, R. Chiu, B. Stein, R.J Imbra, H.J. Rahmsdorf, C. Jonat, P. Herrlich, and M. Karin (1987). Phorbol ester-inducible genes contain a common *cis* element recognised by a TPA-modulated *trans*-acting factor. Cell, 49, 729-739.

Anthony-Cahill, S.J., P.A. Benfield, R. Fairman, Z.R. Wasserman, S.L. Brenner, W.F. Stafford, C. Altenbach, W. Hubbell, and W.F. DeGrado (1992). Molecular characterisation of helix-loop-helix peptides. Science, 255, 979-983.

Arai, H., and P.C. Emson (1986). Regional distribution of neuropeptide K and other tachykinins (neurokinin A, neurokinin B, and substance P) in rat central nervous system. Brain Research, 399, 240-249.

Aso, T., H.A. Vasavada, T. Kawaguchi, F.J. Germino, S. Ganguly, S. Kitajima, S.M. Weissman, and Y. Yasukochi (1992). Characterisation of cDNA for the large subunit of the transcription initiation factor TFIIF. Nature, 355, 461-464.

Assa-Munt, N., R.J. Mortshire-Smith, R. Aurora, W. Herr, and P.E. Wright (1993). The solution structure of the Oct-1 POU-specific domain reveals a striking similarity to the bacteriophage λ repressor DNA-binding domain. Cell, 73, 193-205.

Augereau, P., and P. Chambon (1986). The mouse immunoglobulin heavy-chain enhancer: effect on transcription in vitro and binding of proteins present in HeLa and lymphoid B cell extracts. The EMBO Journal, 5, 1791-1797.

Augusti-Tocco, G., and G. Sato (1969). Establishment of functional clonal lines of neurons from mouse neuroblastoma. Proceedings of the National Academy of Sciences. USA., 64, 311-315.

Ayer, D.E., and W.S. Dynan (1988). Simian virus 40 major late promoter: a novel tripartite structure that includes intragenic sequences. Molecular and Cellular Biology, 8, 2021-2033.

Ayer, D.E., L. Kretzner, and R.N. Eisenman (1993). Mad: a heterodimeric partner for Max that antagonises Myc transcriptional activity. Cell, 72, 211-222.

Ayra, S.K., C. Guo, S.J. Josephs, and F. Wong-Staal (1985). *Trans*-activator gene of human T-lymphotropic virus type III (HTLV-III). Science, 229, 69-73.

Bakker, O., and M.G. Parker (1991). CAAT/enhancer binding protein is able to bind to ATF/CRE elements. Nucleic Acids Research, 19, 1213-1217.

Baniahmad, A., S.Y. Tsai, B.W. O'Malley, and M.-J. Tsai (1992). Kindred S thyroid receptor is an active and constitutive silencer and a repressor for thyroid hormone and retinoic acid responses. Proceedings of the National Academy of Sciences. USA., 89, 10633-10637.

Baniahmad, A., I. Ha, D. Reinberg, S. Tsai, M.-J. Tsai, and B.W. O'Malley (1993). Interaction of human thyroid receptor β with transcription factor TFIIB may mediate target gene derepression and activation by thyroid hormone. Proceedings of the National Academy of Sciences. USA., 90, 8832-8836.

Bannon, M.J., J.-M. Lee, P. Giraus, A. Young, H.U. Affolter, and T.I. Bonner (1986). Dopamine antagonist haloperidol decreases substance P, neuropeptide K, and preprotachykinin mRNAs in rat striatonigral neurons. Journal of Biological Chemistry, 261, 6640-6642.

Bannon, M.J., P.J. Elliot, and E.B. Bunney (1987). Striatal tachykinin biosynthesis: regulation of mRNA and peptide levels by dopamine agonists and antagonists. Molecular Brain Research, 3, 31-37.

Bannon, M.J., M.S. Poosch. D.M. Haverstick, A. Mandal, I.C.H. Xue, K. Shibata, and L.J. Dragovic (1992). Preprotachykinin gene expression in the human basal ganglia: characterisation of mRNAs and pre-mRNAs produced by alternate RNA splicing. Molecular Brain Research, 12, 225-231.

Barber, R.P., J.E. Vaughn, J.R. Slemmon, P.M. Salvaterra, E. Roberts, and S.E. Leeman (1979). The origin and distribution and synaptic relationships of substance P axons in rat spinal cord. Journal of Comparative Neurology, 184, 331-352.

Barberis, A., C.W. Muller, S.C. Harrison, and M. Ptashne (1993). Delineation of two functional regions of transcription factor TFIIB. Proceedings of the National Academy of Sciences. USA., 90, 5628-5632.

Bar-Shavit, Z., R. Goldman, Y. Stabinsky, P. Gottleib, M. Fridkin, V.I. Teichbert, and S. Blumberg (1980). Enhancement of phagocytosis- a newly found activity of substance P residing in its N-terminal tetrapeptide sequence. Biochemical and Biophysical Research Communications, 94, 1445.

Bartel, D.P., M. Sheng, L.F. Lau, and M.E. Greenberg (1989). Growth factors and membrane depolarization activate distinct programs of early response gene expression: dissociation of *fos* and *jun* induction. Genes and Development, 3, 304-313.

Bartolomei, M.S., N.F. Halden, C.R. Cullen, and J.L. Corden (1988). Genetic analysis of the repetitive carboxy-terminal domain of the largest subunit of mouse RNA polymerase II. Molecular and Cellular Biology, 8, 330-339.

Baumruker,T., R. Strum, and W. Herr (1988). OBP100 binds remakably degenerate octamer motifs through specific interactions with flanking sequences. Genes and Development, 2, 1400-1413.

Beato, M. (1989). Gene regulation by steroid hormones. Cell, 56, 335-344.

Beckmann, H., L.K. Su, and T. Kadesch (1990). TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer μE3 motif. Genes and Development, 4, 167-179.

Begley, C.G., S. Lipkowitz, V. Gobel, K.A. Mahon, V. Bertness, A.R. Green, N.M. Gough, and I.R. Kirsch (1992). Molecular characterisation of NSCL, a gene encoding a helix-loop-helix protein expressed in the developing nervous system. Proceedings of the National Academy of Sciences. USA., 89, 38-42.

Benezra, R., R.L. Davis, D. Lockshon, D.L. Turner, and H. Weintraub (1990). The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell, 61, 49-59.

Bengal, E., L. Ransone, R. Scharfmann, V.J. Dwarki, S.J, Tapscott, H. Weintraub, and I.M. Verma (1992). Functional antagonism between c-Jun and MyoD proteins: a direct physical association. Cell, 68, 507-519.

Berg, J.M. (1986). Potential metal-binding domains in nucleic acid binding proteins. Science, 232, 485-487.

Berg, J.M. (1988). Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. Proceedings of the National Academy of Sciences. USA., 85, 99-102.

Bianchi, M.E., M. Beltrame, and G. Paonessa (1989). Specific recognition of cruciform DNA by nuclear protein HMG1. Science, 243, 1056-1059.

Bianchi, M.E., L. Falciola, S. Ferrari, and D.M.J. Lilley (1992). The DNA binding site of HMG1 proteins is composed of two similar segments (HMG boxes) both of

which have counterparts on other eukaryotic regulatory proteins. The EMBO Journal, 11, 1055-1063.

Bienenstock, J., M. Tomioka, H. Matsuda, R.H. Stead, G. Quinonez, G.T. Simon, M.D. Coughlin, and J.A. Denberg (1987). The role of mast cells in inflammatory processes: evidence for nerve/mast cell interactions. International Archives of Allergy and Applied Immunology, 82, 238-243.

Bienz, M., and H.R.B. Pelham (1986). Heat shock regulatory elements function as an enhancer in the *Xenopus* hsp70 gene when linked to a heterologous promoter. Cell, 45, 753-760.

Biggin, M.D., S. Bickel, M. Benson, V. Pirotta, and R. Tjian (1988). *Zeste* encodes a sequence-specific transcription factor that activates the Ultrabithorax promoter in vitro. Cell, 53, 713-722.

Billeter, M., Y. Qian, G. Otting, M. Muller, W.J. Gehring, and K. Wuttrich (1990). Determination of the three-dimensional structure of the Antennapedia homeodomain from Drosophila in solution by H1 nuclear magnetic resonance spectroscopy. Journal of Molecular Biology, 214, 183-197.

Bjorklund, A., and O. Lindvall (1984). Dopamine-containing systems in the CNS. Handbook of Chemical Neuroanatomy, 2, 55-122.

Blackwell, T.K., and H. Weintraub (1990). Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science, 250, 1104-1110.

Blackwell, T.K., L. Kretzner, E.M. Blackwood, R.N. Eisenman, and H. Weintraub (1990). Sequence-specific DNA binding by the c-Myc protein. Science, 250, 1149-1151.

Blackwell, T.K., J. Huang, A. Ma, L. Kretzner, F.W. Alt, R.N. Eisenman, and H. Weintraub (1993). Binding of Myc proteins to canonical and noncanonical DNA sequences. Molecular and Cellular Biology, 13, 5216-5224.

Blackwood, E.M., and R.N. Eisenman (1991). Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. Science, 251, 1211-1217.

Blackwood, E.M., B. Luscher, and R.N. Eisenman (1992a). Myc and Max associate in vivo. Genes and Development, 6, 71-80.

Blackwood, E.M., L. Kretzner, and R.N. Eisenman (1992b). Myc and Max function as a nucleoprotein complex. Current Opinions in Genetics and Development, 2, 227-235.

Bodner, M., J.-L. Castrillo, L.E. Theill, T. Deerink, M. Ellisman, and M. Karin (1988). The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein. Cell, 50, 267-275.

Bohmann, D., T.J. Boss, A. Admon, T. Nishimura, P.K. Vogt, and R. Tjian (1987). Human proto-oncogene c-*jun* encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. Science, 238, 1386-1392.

Bonner, T., H.U. Affolter, A.C. Young, and W.S. Young (1987). A cDNA encoding the precursor of the rat neuropeptide neurokinin B. Molecular Brain Research, 2, 243-249.

Boyer, T.G., and A.J. Berk (1993). Functional interaction of adenovirus E1A with holo-TFIID. Genes and Development, 7, 1810-1823.

Braun, T., G. Buschhausen-Denker, E. Bober, E. Tannich, and H.H. Arnold (1989). A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. The EMBO Journal, 8, 701-709.

Brene, S., N. Lindefors, and H. Persson (1992). Midbrain dopamine neurons regulate preprotachykinin-A mRNA expression in the rat forebrain during development. Molecular Brain Research, 14, 13-19.

Brennan, R.G., S.L. Roderick, Y. Takeda, and B.W. Matthews (1990). Protein-DNA conformational changes in the crystal structure of a λ Cro-operator complex. Proceedings of the National Academy of Sciences. USA., 87, 8165-8169.

Bresnick, E.H., M. Bustin, V. Marsaud, H. Richard-Foy, and G.L. Hager (1992). The transcriptionally active MMTV promoter is depleted of histone H1. Nucleic Acids Research, 20, 273-278.

Briggs, M.R., J.T. Kadonga, S.P. Bell, and R. Tjian (1986). Purification and biochemical characterisation of the promoter-specific transcription factor Sp1. Science, 234, 47-52.

Brimijoin, S., J.M. Lundberg, E. Brodin, T. Hokfelt, and G. Nilsson (1980). Axonal transport of substance P in the vagus and sciatic nerve of the guinea-pig. Brain Research, 191, 443-457.

Brodeur, G.M., and M.N. Goldstein (1976). Histochemical demonstration of an increase in acetylcholinesterase in established lines of human and mouse neuroblastomas by nerve growth factor. Cytobios, 16, 133-138.

Brou, C., S. Chaudary, I. Davidson, Y. Lutz, J. Wu, J.-M. Egly, L. Tora, and P. Chambon (1993). Distinct TFIID complexes mediate the effect of different transcriptional activators. The EMBO Journal, 12, 489-499.

Brown, R.S., C. Sander, and P. Argos (1985). The primary structure of transcription factor TFIIIA has twelve consecutive repeats. FEBS Letters, 186, 271-274.

Bruggemeier, U., M. Kalff, S. Franke, C. Scheidereit, and M. Beato (1991). Ubiquitous transcription factor OTF-1 mediates induction of the MMTV promoter through synergistic activation with hormone receptors. Cell, 64, 565-572.

Buck, S.H., and E. Burcher (1986). The tachykinins: a family of peptides with a brood of 'receptors'. Trends in Pharmacological Sciences, 7, 65-68.

Buettner, R., P. Kannan, A. Imhof, R. Bauer, S.O. Yim, R. Glockshuber, M.W. VanDyke, and M.A. Tainsky (1993). An alternatively spliced mRNA from the AP-2 gene encodes a negative regulator of transcriptional activation by AP-2. Molecular and Cellular Biology, 13, 4174-4185.

Buratowski, S., S. Hahn, L. Guarente, P.A. Sharp, and L. Guarente (1988). Function of a yeast TATA binding protein in a mammalian transcription system. Nature, 334, 37-42.

Buratowski, S., S. Hahn, L. Guarente, and P.A. Sharp (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell, 56, 549-561.

Buratowski, S., M. Spota, J. Greenblatt, and P.A. Sharp (1991). RNA polymerase II-associated proteins are required for a DNA conformation change in the transcriptional initiation complex. Proceedings of the National Academy of Sciences. USA., 88, 7509-7513.

Buratowski, S., and H. Zhou (1992). Transcription factor IID mutants defective for interaction with transcription factor IIA. Science, 255, 1130-1132.

Buratowski, S., and H. Zhou (1993). Functional domains of transcription factor TFIIB. Proceedings of the National Academy of Sciences. USA., 90, 5633-5637.

Bury, R.W., and M.L. Mashford (1976). Biological activity of C-terminal partial sequences of substance P. Journal of Medicinal Chemistry, 19, 854-856.

Cabrera, C.V., A. Martinez-Arias, and M. Bate (1987). The expression of three members of the *achaete-scute* gene complex correlates with neuroblast segregation in drosophila. Cell, 50, 425-433.

Cadena, D.L., and M.E. Dahmus (1987). Messenger mRNA synthesis in mammalian cells is catalysed by the phosphorylated from of RNA polymerase II. Journal of Biological Chemistry, 262, 12468-12474.

Cao, Z., R.M. Umek, and S.L. McKnight (1991). Regulated expression of the three C/EBP isoforms during adipose conversion of 3T3-L1 cells. Genes and Development, 5, 1538-1552.

Carr, C.S., and P.A. Sharp (1990). A helix-loop-helix protein related to the immunoglobulin E Box-binding proteins. Molecular and Cellular Biology, 10, 4384-4388.

Carter, M.S., and J.E. Krause, (1990). Structure expression and some regulatory mechanisms of the rat preprotachykinin gene encoding substance P, neurokinin A, neuropeptide K and neuropeptide γ. Journal of Neuroscience, 10, 2203-2214.

Carthew, R.W., L.A. Chodosh, and P.A. Sharp (1985). An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell, 43, 439-448.

Cavallini, B., I. Faus, H. Matthes, J.M. Chipoulet, B. Winsor, J.M. Egly, and P. Chambon (1989). Cloning of the gene encoding the yeast protein BTF11, which can substitute for the human TATA box binding factor. Proceedings of the National Academy of Sciences. USA., 86, 9803-9807.

Ceska, T.A., M. Lamers, P. Monaci, A. Nicosia, R. Cortese, and D. Suck (1993). The X-ray structure of an atypical homeodomain present in the rat liver transcription factor LFB/HNF1 and implications for DNA binding. The EMBO Journal, 12, 1805-1810.

Chalfie, M., H.R. Horvitz, and J.E Sulston (1981). Mutations that lead to reiterations in the cell lineages of *C. elegans*. Cell, 24, 583-586.

Chang, M.M., S.E. Leeman, and H.D. Niall (1971). Amino acid sequence of substance P. Nature, 232, 86-87.

Chao, M.V. (1992). Growth factor signalling: where is the specificity?. Cell, 68, 995-997.

Chapman, K., V. Lyons, and A.J. Harmar (1993). The sequence of 5' flanking DNA from the rat preprotachykinin gene; analysis of putative transcription factor binding sites. Biochimica et Biophysica Acta, 1172, 361-363.

Chavier, P., M. Zerial, P. Lemair, J. Almendral, R. Bravo, and P. Charnay (1988). A gene encoding a protein with Zn fingers is activated during G0/G1 transition in cultured cells. The EMBO Journal, 7, 29-35.

Cho, K.W.Y., K. Khalili, R. Zandomeni, and R. Weinmann (1985). The gene encoding the largest subunit of RNA polymerase II. Journal of Biological Chemistry, 260, 15204-15210.

Choy, B., and M.R. Green (1993). Eukaryotic activators function during multiple steps of preinitiation complex assembly. Nature, 366, 531-536.

Christy, B.A., L.F. Lau, and D. Nathans (1988). A gene activated in mouse 3T3 cells by serum growth factors encodes a protein with "zinc finger" sequences. Proceedings of the National Academy of Sciences. USA., 85, 7857-7861.

Christy, R.J., V.W. Yang, J.M. Ntambi, D.E. Geiman, W.H. Landschulz, A.D. Friedman, Y. Nakabeppu, T.J. Kelly, and M.D. Lane (1989). Differentiation-induced gene expression in 3T3-L1 preadipocytes: CCAAT/enhancer binding protein interacts with and activates the promoters of two adipocyte-specific genes. Genes and Development, 3, 1323-1335.

Christy, B.A., L.K. Sanders, L.F. Lau, N.G. Copeland, N.A. Jenkins, and D. Nathans (1991). An Id-related helix-loop-helix protein encoded by a growth factor-inducible gene. Proceedings of the National Academy of Sciences. USA., 88, 1815-1819.

Clark, S.P., C.D. Lewis, and G. Felsenfeld (1990). Properties of BGP1, a poly(dG)-binding protein from chicken erythrocytes. Nucleic Acids Research, 18, 5119-5126.

Clark, K.L., E.D. Halay, E. Lai, and S.K. Burley (1993). Co-crystal structure of the HNF-3/*fork head* DNA-recognition motif resembles histone H5. Nature, 364, 412-420.

Clarke, N.D., L.J. Beamer, H.R. Goldberg, C. Berkower, and C.O. Pabo (1991). The DNA binding arm of λ repressor: critical contacts from a flexible region. Science, 254, 267-270.

Clerc, R.G., L.M. Corcoran, J.H. LeBowitz, D. Baltimore, and P.A. Sharp (1988). The B-cell-specific Oct-2 protein contains POU-box and homeo-box-type domains. Genes and Development, 2, 1570-1581.

Cohen, D.R., and T. Curran (1988). *fra*-1: A serum inducible, cellular immediate-early gene that encodes a Fos-related antigen. Molecular and Cellular Biology, 8, 2063-2069.

Cohen, D.R., P.C.P. Ferreira, R. Gentz, B.R. Franza Jr., and T. Curran (1989). The product of a *fos*-related gene, *fra*-1, binds co-operatively to the AP1 site with Jun: transcription factor AP1 is comprised of multiple protein complexes. Genes and Development, 3, 173-184.

Colgan, J., S. Wampler, and J.L. Manley (1993). Interaction between a transcriptional activator and transcripton factor IIB *in vivo*. Nature, 362, 549-553.

Comai, L., N. Tanese, and R. Tjian (1992). The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1. Cell, 68, 965-976.

Comb, M., N.C Birnberg, A. Seasholtz, E. Herbert, and H. Goodman (1986). A cyclic AMP- and phorbol ester- inducible DNA element. Nature, 323, 353-356.

Conaway, J.W., M.W. Bond, and R.C. Conaway (1987). An RNA polymerase II transcription system from rat liver. Journal of Biological Chemistry, 262, 8293-8297.

Conaway, J.W., and R.C. Conaway (1989a). A mutisubunit transcription factor essential for accurate transcription by RNA polymerase II. Journal of Biological Chemistry, 264, 2357-2362.

Conaway, R.C., and J.W. Conaway (1989b). An RNA polymerase II transcription factor has an associated DNA-dependent ATPase (dATPase) activity strongly stimulated by the TATA region of promoters. Proceedings of the National Academy of Sciences. USA., 86, 7356-7360.

Conaway, J.W., D. Reines, and R.C. Conaway (1990a). Transcription initiated by RNA polymerase II and purified transcription factors from liver. Journal of Biological Chemistry, 265, 7552-7558.

Conaway, J.W., E. Travis, and R.C. Conaway (1990b). Transcription initiated by RNA polymerase II and purified transcription factors from liver. Journal of Biological Chemistry, 265, 7564-7569.

Conaway, J.W., J.P. Hanley, K.P. Garret, and R.C. Conaway (1991). Transcription initiated by RNA polymerase II and purified transcription factors from liver. Journal of Biological Chemistry, 266, 7804-7811.

Conaway, R.C., J.N. Bradsher, and J.W. Conaway (1992a). Mechanism of assembly of the RNA polymerase II preinitiation complex. Journal of Biological Chemistry, 267, 8464-8467.

Conaway, J.W., J.N. Bradsher, and Conaway (1992b). Mechanism of assembly of the RNA polymerase II preinitiation complex. Journal of Biological Chemistry, 267, 10142-10148.

Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon (1980). Promoter sequences of eukaryotic protein-encoding genes. Science, 209, 1405-1414.

Corden, J.L., D.L. Cadena, J.M. Ahearn Jr., and M.E. Dahmus (1985). A unique structure at the carboxy-terminus of the largest subunit of eukaryotic RNA polymerase II. Proceedings of the National Academy of Sciences. USA., 82, 7934-7938.

Cormack, B.P., M. Strubin, A.S. Ponticelli, and K. Struhl (1991). Functional differences between yeast and human TFIID are localised to the highly conserved region. Cell, 65, 341-348.

Cortes, P., O. Flores, and D. Reinberg (1992). Factors involved in specific transcription by mammalian RNA polymerase II: purification and analysis of transcription factor IIA and identification of transcription factor IIJ. Molecular and Cellular Biology, 12, 413-421.

Coulombe, B., M. Killeen, P. Liljelund, B. Honda, H. Xiao, C.J. Injles, and J. Greenblatt (1992). Identification of three mammalian proteins that bind to the yeast TATA box protein TFIID. Gene Exp., 2, 99-109.

Courery, A.J., and R. Tjian (1988). Analysis of Sp1 in vivo reveals multiple transcriptonal domains, including a novel glutamine-rich activation motif. Cell, 55, 887-898.

Couture, R., A. Fournier, J. Magnan, S. St.-Pierre, and D. Regoli (1979). Structure-activity studies on substance P. Canadian Journal of Physiological Pharmacology, 57, 1427-1436.

Cress, W.D., and S.J. Triezenberg (1991). Critical structural elements of the VP16 transcriptional activation domain. Science, 251, 87-90.

Crossley, M., and G.G. Brownlee (1990). Disruption of a C/EBP binding site in the factor IX promoter is associated with haemophilia B. Nature, 345, 444-446.

Curran, T., and N.M. Teich (1982). Candidate product of the FBJ murine osteosarcoma virus oncogene. Journal of Virology, 42, 114-122

Curran, T., A.D. Miller, L. Zokas, and I.M. Verma (1984). Viral and cellular *fos* proteins: a comparative analysis. Cell, 36, 259-268.

Curran, T., and J. Morgan (1985). Superinduction of *c-fos* by nerve growth factor in the presence of peripherally active benzodiazepines. Science, 229, 1265-1268.

Curran, T., C. Van Beveren, N. Ling, and I.M. Verma (1985). Viral and cellular *fos* proteins are complexed with a 39,000-dalton cellular protein. Molecular and Cellular Biology, 5, 167-172.

Curran, T., and B.R. Faranza, Jr. (1988). Fos and Jun: the AP-1 connection. Cell, 55, 395-397.

Dahmus, M.E., and C. Kedinger (1983). Transcription of the adenovirus-2 major late promoter by monoclonal antibody directed against RNA polymerase IIo and IIA. Journal of Biological Chemistry, 258, 2303-2307.

Dalsgaard, C.-J., C.E. Jonsson, T. Hokfelt, and A.C. Cuello (1983). Localisation of substance P-immunoreactive nerve fibers in the human digital skin. Experientia, 39, 1018-1020.

Dalsgaard, C.-J., A. Haegerstrand, E. Brodin, E. Theodorsson-Norheim, and T. Hokfelt (1985). Neurokinin A-like immunoreactivity in rat primary sensory neurons: coexistence with substance P. Histochemistry, 83, 407-414.

Dalsgaard, C.-J., J. Jernbeck, W. Stains, J. Kjartansson, A. Haegerstrand, T. Hokfelt, E. Brodin, A.C. Cuello, and J.C. Brown (1988). Calcitonin gene-related peptide like immunoreactivity in nerve fibers in the human skin: relation to fibers containing substance P-, somatostatin- and vasoactive intestinal polypeptide-like immunoreactivity. Histochemistry, 91, 35-38.

Das, G., and W. Herr (1993). Enhanced activation of the human histone H2B promoter by an Oct-1 variant generated by alternative splicing. Journal of Biological Chemistry, 268, 25026-25032.

Dash, P.K., K.A. Karl, M.A. Colicos, R. Prywes, and E.R. Kandel (1991). cAMP response element-binding protein is activated by Ca^{2+}/calmodulin- as well as cAMP-dependent protein kinase. Proceedings of the National Academy of Sciences. USA., 88, 5061-5065.

Davis, R.L., H. Weintraub, and A.B. Lassar (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell, 51, 987-1000.

Davis, T.L., A.B. Firulli, and A.J. Kinniburgh (1989). Ribonucleoprotein and protein factors bind to an H-DNA-forming *c-myc* DNA element: possible regulators of the *c-myc* gene. Proceedings of the National Academy of Sciences. USA., 86, 9682-9686.

Davis, R.L., P.F. Cheng, A.B. Lassar, and H. Weintraub (1990). The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. Cell, 60, 733-746.

Davison, B.L., J. Egly, E.R. Mulvihill, and P. Chambon (1983). Formation of stable preinitiation complexes between eukaryotic class B transcription factors and promoter sequences. Nature, 301, 680-686.

DeJong, J., and R.G. Roeder (1993). A single cDNA, hTFIIA/α, encodes both the p35 and p19 subunits of human TFIIA. Genes and Development, 7, 2220-2234.

Dekker, N., M. Cox, R. Boelens, C.P. Verrijzer, P.C. van der Vliet, and R. Kaptein (1993). Solution structure of the POU-specific DNA-binding domain of Oct-1. Nature, 362, 852-854.

DePinho, R.A., K.S. Hatton, A. Tesfaye, G.D. Yancopoulos, and F.W. Alt (1987). The human *myc* gene family: structure and activity of L-*myc* and an L-*myc* pseudogene. Genes and Development, 1, 1311-1326.

DesJardins, E., and N. Hay (1993). Repeated CT elements bound by zinc finger proteins control the absolute and relative activities of the two principal human *c-myc* promoters. Molecular and Cellular Biology, 13, 5710-7724.

Dignam, J.D., P.L. Martin, B.S. Shastry, and R.G. Roeder (1983). Eukaryotic gene transcription with purified components. Methods In Enzymology, 101, 582-598.

Donaldson, L., and J.P. Capone (1992). Purification and characterization of the carboxyl-terminal transactivation domain of Vmw65 from herpes simplex virus type 1. Journal of Biological Chemistry, 267, 1411-1414.

Donaldson, L.F., A.J. Harmar, D.S. McQueen, and J.R. Seck l (1992). Increased expression of preprotachykinin, calcitonin gene-related peptide, but not vasoactive intestinal peptide messenger RNA in dorsal root ganglia during the development of adjuvant monoarthritis in the rat. Molecular Brain Research, 16, 143-149.

Donaldson, L.F., D.S. McQueen, and J.R. Seckl (1993). Expression of mRNA encoding the transcription factor AP2 increases in dorsal root ganglion neurons in response to adjuvant injection in the rat in vivo. Proceedings of The Physiological Society, C165, 15.

Du, H., A.L. Roy, and R. Roeder (1993). Human transcription factor USF stimulates transcription through the initiator elements of the HIV-1 and the Ad-ML promoters. The EMBO Journal, 12, 501-511.

Durrin, L.K., R.K. Mann, and M. Grunstein (1992). Nucleosome loss activates *CUP 1* and *HIS 1* promoters to fully induced levels in the yeast *saaharomyces cerevisae.* Molecular and Cellular Biology, 12, 1621-1629.

Dynan, W.S., and R. Tjian (1983a). Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell, 32, 669-680.

Dynan, W.S., and R. Tjian (1983b). The promoter-specific transcripton factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell, 35, 79-87.

Dynan, W.S., and R. Tjian (1985). Control of eukaryotic messanger RNA synthesis by sequence-specific DNA-binding proteins. Nature, 316, 774-778.

Dynan, W.S., J.D. Saffer, W.S. Lee, and R. Tjian (1985). Transcription factor Sp1 recognizes promoter sequences from the monkey genome that are similar to the simian virus 40 promoter. Proceedings of the National Academy of Sciences. USA., 82, 4915-4919.

Dynan, W.S., S. Sazer, R. Tjian, and R.T. Schimke (1986). Transcription factor Sp1 recognizes a DNA sequence in the mouse dihydrofolate reductase promoter. Nature, 319, 246-248.

Dynlacht, B.D., T. Hoey, and R. Tjian. (1991). Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. Cell, 66, 563-576.

Edlund, T., M.D. Walker, P.J. Barr, and W.J. Rutter (1985). Cell specific expression of the rat insulin gene: evidence for the role of two distinct 5' flanking sequences. Science, 230, 912-916.

Egly, J.M., N.G. Miyamoto, V. Moncollin, and P. Chambon (1984). Is actin a transcription initiation factor for RNA polymerase B? The EMBO Journal, 3, 2363-2371.

Eisenmann, D.M., C. Dollard, and F. Winston (1989). *SPT15*, the gene encoding the yeast TATA binding factor TFIID is required for normal transcription initiation *in vivo.* Cell, 58, 1183-1191.

Elgin, S.C.R. (1984). Anatomy of hypersensitive sites. Nature, 209, 213-214.

Elsholtz, H.P., V.R. Albert, M.N. Treacy, and M.G. Rosenfeld (1990). A two base change in a POU factor-binding site switches pituitary-specific to lymphoid-specific gene expression. Genes and Development, 4, 43-51.

Erspamer, V. (1971). Biogenic amines and active polypeptides of the amphibian skin. Annual Review of Pharmacology, 11, 327-350.

Falkner, F.G., and H.G. Zachau (1984). Correct transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. Nature, 310, 71-74.

Farmer, G.E., J. Bargonetti, H. Zu, P.N. Friedman, R. Pryers, and C. Prives (1992). Wild-type p53 activates transcription *in vitro*. Nature, 358, 83-86.

Feldmann, T., R. Alex, J. Suckow, R. Dildrop, B. Kisters-Woike, and B. Muller-Hill (1993). Single exchanges of amino acids in the basic region change the specificity of N-Myc. Nucleic Acids Research, 21, 5050-5058.

Felten, D.L., S.Y. Felten, S.L. Carlson, J.A. Olschowka, and S. Livant (1985). Noradrenergic and peptidergic innervation of lymphoid tissue. Journal of Immunology, 135, 755s-765s.

Feng, S., and E.C. Holland (1988). HIV-1 *tat trans*-activation requires the loop sequence within *tar*. Nature, 334, 165-167.

Ferre-D'Amare, A., G.C. Pendergast, E.B. Ziff, and S.K. Burley (1993). Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature, 363, 38-44.

Fink, T., and E. Weihe (1988). Multiple neuropeptides in nerves supplying mammalian lymph nodes: messenger candidates for sensory and autonomic neuroimmunomodulation. Neuroscience Letters, 90, 39-44.

Finkelstein, A., C.F. Kostrub, J. Li, D.P. Chavez, B.Q. Wang, S.M. Fang, J. Greenblatt, and Z.F. Burton (1992). A cDNA encoding RAP74, a general initiation factor for transcription by RNA polymerase II. Nature, 355, 464-467.

Finney, M., G. Ruvkin, and H.R. Horvitz (1988). The C. elegans cell lineage and differentiation gene unc-86 encodes a protein containing a homeodomain and extended sequence similarity to mammalian transcription factors. Cell, 55, 757-769.

Fire, A., M. Samuels, and P.A. Sharp (1984). Interactions between RNA polymerase II, factors, and template leading to accurate transcription. Journal of Biological Chemistry, 259, 2509-2516.

Fischer, L., M. Gerard, C. Chalut, Y. Lutz, S. Humbert (1992). Cloning of the 62-kilodalton component of basic transcription factor BTF2. Science, 257, 1392-1395.

Fisher, D.E., C.S. Carr, L.A. Parent, and P.A. Sharp (1991). TFEB has DNA-binding and oligomerization properties of a unique helix-loop-helix/leucine zipper family. Genes and Development, 5, 2342-2352.

Fisher, F., and C.R. Goding (1992). Single amino acid substitutions alter helix-loop-helix protein specificity for bases flanking the core CANNTG motif. The EMBO Journal, 11, 4103-4109.

Fisher, F., D.H. Crouch, P.-S. Jayaraman, W. Clark, D.A.F. Gillespie, and C.R. Goding (1993). Transcription activation by Myc and Max: flanking sequences target activation to a subset of CACGTG motifs *in vivo*. The EMBO Journal, 12, 5075-5082.

Fitzgerald, M., P.D. Wall, M. Goedert, and P.C. Emson (1985). Nerve growth factor counteracts the neurophysiological and neurochemical effects of chronic sciatic nerve section. Brain Research, 332, 131-141.

Flores, O., E. Maldonada, and D. Reinberg (1989). Factors involved in specific transcription by RNA polymerase II. Journal of Biological Chemistry, 264, 8913-8921.

Flores, O., I. Ha, and D. Reinberg (1990). Factors involved in specific transcription by mammalian RNA polymerase II. Journal of Biological Chemistry, 263, 5629-5634.

Flores, O., H. Lu, M. Killeen, J. Greenblatt, Z.F. Burton, and D. Reinberg (1991). The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. Proceedings of the National Academy of Sciences. USA., 88, 9999-10003.

Flores, O., H. Lu, and D. Reinberg (1992). Factors involved in specific transcription by mammalian RNA polymerase II. Journal of Biological Chemistry, 267, 2786-2793.

Foulkes, N.S., E. Borrelli, and P. Sassone-Corsi (1991). CREM gene: use of alternative DNA-binding domains generates multiple antagonists of cAMP-induced transcription. Cell, 64, 739-749.

Frankel, A.D., and C.O. Pabo (1988). Fingering too many proteins. Cell, 53, 675.

Freedman, L.P., B.F. Luisi, Z.R. Korszun, R. Basavappa, P.B. Sigler, and K.P. Yamamoto (1988). The function and structure of the metal coordination sites within the glucocorticoid receptor DNA binding domain. Nature, 334, 543-546.

Freidin, M., and J.A. Kessler (1991). Cytokine regulation of substance P expression in sympathetic neurons. Proceedings of the National Academy of Sciences. USA., 88, 3200-3203.

Friedman, A.D., W.H. Landschultz, and S.L. McKnight (1989). CCAAT/enhancer binding protein activates the promoter of the serum albumin gene in cultured hepatoma cells. Genes and Development, 3, 1314-1322.

Furness, J.B., R.E. Papka, N.G. Della, M. Costa, and R.L. Eskay (1982). Substance P-like immunoreactiity in nerves associated with the vascular system of guinea-pigs. Neuroscience, 7, 447-459.

Garrett, K.P., H. Serizawa, J.P. Hanley, J.N. Bradsher, A. Tsuboi, N. Arai, T. Yokota, K. Arai, R.C. Conaway, and J.W. Conaway (1992). The carboxyl terminus of RAP 30 is similar in sequence to region 4 of bacterial sigma factors. Journal of Biological Chemistry, 267, 23942-23949.

Gasparovic, I., S. Hadzovic, S. Hukovic, and P. Stein (1964). Contribution to the theory that substance P has a transmitter role in sensitive pathways. Medical Experiments, 10, 303-306.

Geiser, A.G., K.J. Busman, S.J. Kim, R. Lafyatis, M.A. O'Reilly, R. Webbink, A.B. Roberts, and M.B. Sporn (1993). Regulation of the transforming growth factor-beta 1 and -beta 3 promoters by the transcription factor Sp1. Gene, 129, 223-228.

Gerard, M., L. Fischer, V. Moncollin, J.M. Chipoulet, P. Chambon, and J.M. Egly (1991). Purification and interaction properties of the human RNA polymerase B (II) general transcription factor BTF2. Journal of Biological Chemistry, 266, 20940-20945.

Giese, K., J. Cox, and R. Grosschedl (1992). The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. Cell, 69, 185-195.

Gilchrist, C.A., C.F. Morrison, K.E. Chapman, and A.J. Harmar (1991). Identification of nerve growth factor responsive sequences within the 5' region of the bovine preprotachykinin gene. DNA and Cell Biology, 10, 743-749.

Gill, G., J.R. Smith, J.L. Goldstein, C.A. Slaughter, K. Orth, M.S. Brown, and T.F. Osborne (1988). Multiple genes encode nuclear factor 1-like proteins that bind to the promoter for 3-hydroxy-3-methylglutaryl-coenzyme A reductase. Proceedings of the National Academy of Sciences. USA., 85, 8963-8967.

Gill, G., and R. Tjian (1991). A highly conserved domain of TFIID displays species specificty in vivo. Cell, 65, 333-340.

Giniger, E., and M. Ptashne (1987). Transcription in yeast activated by a putative amphipathic α helix linked to a DNA binding unit. Nature, 330, 670-673.

Ginsberg, D., F. Mechta, M. Yaniv, and M. Oren (1991). Wild-type p53 can down-modulate the activity of various promoters. Proceedings of the National Academy of Sciences. USA., 88, 9979-9983.

Gomez-Cuadrado, A., S. Rousseau, J. Renaud, and A. Ruiz-Carrillo (1992). Repression of the H5 histone gene by a factor from erythrocytes that binds to the region of transcription initiation. The EMBO Journal, 11, 1857-1866.

Goodwin, G.H., C. Sanders, and E. Johns (1973). A new group of chromatin-associated proteins with a high content of acidic and basic amino acids. European Journal of Biochemistry, 38, 14-19.

Goutte, C., and A.D. Johnson (1988). a1 protein alters the DNA binding specificty of α2 repressor. Cell, 52, 875-882.

Greenberg, M.E., L.A. Greene, and E.B. Kiff (1985). Nerve growth factor and epidermal growth factor induce rapid transient changes in proto-oncogene transcription factors in PC12 cells. Journal of Biological Chemistry, 260, 14101-14110.

Greenblatt, J. (1991a). Roles of TFIID in transcriptional initiation by RNA polymerase II. Cell, 66, 1067-1070.

Greenblatt, J. (1991b). RNA polymerase-associated transcription factors. Trends in Biochemical Science, 16, 408-411.

Greene, L.A, and E.M. Shooter (1980). The nerve growth factor: biochemistry, synthesis, and mechanism of action. Annual Review of Neuroscience, 3, 353-402.

Gregor, P.D., M. Sawadogo, and R.G. Roeder (1990). The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. Genes and Development, 4, 1730-1740.

Gronemeyer, H., B. Turcotte, C. Quirin-Stricker, M.T. Bocquel, M.E. Meyer, Z. Krozowski, J.M. Jeitsch, T. Lerouge, J.M. Garnier, and P. Chambon (1987). The chicken progesterone receptor: sequence, expression and functional analysis. The EMBO Journal, 6, 3985-3994.

Grosschedl., R., and M.L. Birnsteil (1980). Identification of regulatory sequences in the prelude sequences of an H2 histone gene by the study of specific deletion mutants in vitro. Proceedings of the National Academy of Sciences. USA., 77, 1432-1436.

Guarente, L. (1987). Regulatory proteins in yeast. Annual Review of Genetics. 21, 425-452.

Guillemot, F., L. Li-Ching, J.E. Johnson, A. Auerbach, D.J. Anderson, and A.L. Joyner (1993). Mammalian achaete-scute homologue 1 is required for the early development of olfactory and autonomic neurons. Cell, 75, 463-476.

Ha, I., W.S. Lane, and D. Reinberg (1991). Cloning of a human gene encoding the general initiation factor IIB. Nature, 352, 689-695.

Habener, J.F. (1990). Cyclic AMP response element binding proteins: a cornucopia of transcription factors. Molecular Endocrinology, 4, 1087-1094.

Hahn, S., S. Buratowski, P.A. Sharp, and L. Guarente (1989a). Isolation of the gene encoding the yeast TATA binding protein TFIID: a gene identified to the *SPT15* suppressor of Ty element insertions. Cell, 58, 1173-1181.

Hahn, S., S. Buratowski, P.A. Sharp, and L. Guarente (1989b). Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. Proceedings of the National Academy of Sciences. USA., 86, 5718-5722.

Hahn, S., S. Buratowski, P.A. Sharp, and L. Guarente (1989c). Identification of a yeast protein homologous in function to the mammalian general transcripton factor, TFIIA. The EMBO Journal, 8, 3379-3382.

Hahn, S. (1993). Structure(?) and function of acidic transcription activators. Cell, 72, 481-483.

Hai, T., F. Liu, W.J. Coukos, and M.R. Green (1989). Transcription factor ATF cDNA clones: an extensive family of leucine zipper proteins able to selectively form DNA-binding heterodimers. Genes and Development, 3, 2083-2090.

Hai, T and T. Curran (1991). Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. Proceedings of the National Academy of Sciences. USA., 88, 3720-3724.

Halazonetis, T.D., K. Georgopoulos, M.E. Greenberg, and P. Leder (1988). c-Jun dimerises with itself and with c-Fos, forming complexes of different DNA affinities. Cell, 55, 917-924.

Halazonetis, T.D., and A.N. Kandil (1992). Predicted structural similarities of the DNA binding domains of c-Myc and endonuclease EcoR1. Science, 255, 464-467.

Hanvey, J.C. , M. Schimizu, and R.D. Wells (1988). Intramolecular DNA triplexes in supercoiled plasmids. Proceedings of the National Academy of Sciences. USA., 85, 6292-6296.

Hara, Y., A.C. Rovescalli, Y. Kim, and M. Nirenberg (1992). Structure and evolution of four POU domain genes expressed in the mouse brain. Proceedings of the National Academy of Sciences. USA., 89, 3280-3284.

Hard, T., E. Kellenbach, R. Boelens, B.A. Maler, K. Dahlman, K.R. Yamamoto, L.P. Freedman,I. Carlstedt-Duke, J.-A. Gustafsson, and R. Kaptein (1990). Solution structure of the glucacorticoid receptor DNA-binding domain. Science, 249, 157-160.

Harmar, A.J., and P. Keen (1982). Synthesis, and central, and peripheral axonal transport of substance P in a dorsal root ganglion-nerve preparation in vitro. Brain Research, 231, 379-386.

Harmar, A.J., V. Hyde, and K. Chapman (1990). Identification and cDNA sequence of δ-preprotachykinin, a fourth splicing variant of the rat substance P precursor. FEBS Letters, 275, 22-24.

Hart, R.P., A.M. Shadiack, and G.M. Jonakait (1991). Substance P gene expression is regulated by interleukin-1 in cultured sympathetic ganglia. Journal of Neuroscience Research, 29, 282-291.

Harvey, R.P., A.J. Robins, and J.R.E. Wells (1982). Independently evolving chicken histone H2B genes: identification of a ubiquitous H2B-specific 5' element. Nucleic Acids Research, 10, 7851-7863.

Haslinger, A., and Karin, M. (1985). Upstream promoter elements of the human metallothionein-IIa gene can act like an enhancer element. Proceedings of the National Academy of Sciences. USA., 82, 8572-8576.

Hay, N., J.M. Bishop, and D. Levens (1987). Regulatory elements that modulate expression of human c-myc. Genes and Development, 1, 659-671.

He, X., M.N. Treacy, D.M. Simmons, H.A. Ingraham, L.W. Swanson, and M.G. Rosenfeld (1989). Expression of a large family of POU-domain regulatory genes in mammalian brain development. Nature, 340, 35-42.

Helmann, J.D., and M.J. Chamberlin (1988). Structure and function of bacterial sigma factors. Annual Review of Biochemistry, 57, 839-872.

Henken, D.B., W.P. Battisti, M.F. Chesselet, M. Murray, and A. Tessler (1990). Expression of β-preprotachykinin mRNA and tachykinins in rat dorsal root ganglion cells following peripheral or central axotomy. Neuroscience, 39, 733-742.

Henry, J.L. (1976). Effects of substance P on functionally defined units in cat spinal cord. Brain Research, 114, 439-451.

Herr, W., R. Sturm, R. Clerc, L. Corcoran, D. Baltimore, P. Sharp, H. Ingraham, M. Rosenfeld, M. Finney, G. Ruvkun, and H. Horvitz (1988). The POU domain: a large conserved region in the mammalian *pit-1*, *oct-1*, *oct-2*, and *Caenorhabditis elegans unc-86* gene products. Genes and Development, 2, 1513-1516.

Hirai, S., R. Ryseck, F. Mechta, R. Bravo, and M. Yanin (1989). Characterisation of *jun*D: a new member of the *jun* proto-oncogene family. The EMBO Journal, 8, 1433-1439.

Hisatake, K., S. Hasegawa, R. Takada, Y. Nakatani, M. Horikoshi, and R.G. Roeder (1993a). The p250 subunit of native TATA box-binding factor TFIID is the cell-cycle regulatory protein CCG1. Nature, 362, 179-181.

Hisatake, K., R.G. Roeder, and M. Horikoshi (1993b). Functional dissection of TFIIB domains required for TFIIB-TFIID-promoter complex formation and basal transcription activity. Nature, 363, 744-747.

Hoejimakers, J.H.J. (1990). Cryptic initiation sequence revealed. Nature, 343, 417-418.

Hoey, T., B.D. Dynlacht, M.G. Peterson, B.F. Pugh, and R. Tjian (1990). Isolation and characterisation of the *drosophila* gene encoding the TATA box binding protein. Cell, 61, 1179-1186.

Hoffmann, A., E. Sinn, T. Yamamoto, J. Wang, A. Roy, M. Horikoshi, and R.G. Roeder (1990). Highly conserved core domain and unique N terminus with presumptive regulatory motifs in a human TATA factor (TFIID). Nature, 346, 387-390.

Hokfelt, T., J.-O. Kellerth, G. Nilsson, and B. Pernow (1975). Substance P: localization in central nervous system and in some primary sensory neurons. Science, 190, 889-890.

Hoopes, B.C., J.F. LeBlanc, and D.K. Hawley (1992). Kinetic analysis of yeast TFIID-TATA box complex formation suggests a multi-step pathway. Journal of Biological Chemistry, 267, 11539-11547.

Hope, I.A., and K. Struhl (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell, 46, 885-894.

Hope, I.A., S. Mahadevan, and K. Struhl (1988). Structural and functional characteristion of the short acidic transcriptional activation region of yeast GCN4 protein. Nature, 333, 635-640.

Horikoshi, M., T. Hai, Y.-S. Lin, M.R. Green, and R.G. Roeder (1988). Transcription factor ATF interacts with the TATA factor to facilitate establishment of a preinitiaion complex. Cell, 54, 1033-1042.

Horikoshi, M., C.K. Wang, H. Fujii, J.A. Cromlich, B.A. Weil, and R.G. Roeder (1989). Cloning and structure of a yeast gene encoding a general transcription factor TFIID that binds to the TATA box. Nature, 341, 299-303.

Horikoshi, M., T. Yamamoto, Y. Ohkuma, P.A. Weil, and R.G. Roeder (1990). Analysis of structure-function relationships of yeast TATA box binding factor TFIID. Cell, 61, 1171-1178.

Hsu, W., T.K. Kerpolla, P.-L. Chen, T. Curran, and S. Chen-Kiang (1994). Fos and Jun repress transcription activation by NF-IL6 through association at the basic leucine zipper region. Molecular and Cellular Biology, 14, 268-276.

Htun, H., E. Lund, and J.E. Dahlberg (1984). Human U1 RNA genes contain an unusually sensitive nuclease S1 cleavage site within the conserved 3' flanking region. Proceedings of the National Academy of Sciences. USA., 81, 7288-7292.

Htun, H., and J.E. Dahlberg (1988). Single strands, triple strands, and kinks in H-DNA. Science, 241, 1791-1796.

Hu, Y.F., B. Luscher, A. Admon, N. Mermond, and R. Tjian (1990). Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. Genes and Development, 4, 1741-1752.

Hu, J.-S., E.N. Olson, and R.E. Kingston (1992). HEB: a helix-loop-helix protein related to E2A and ITF2 that can modulate the DNA-binding ability of myogenic regulatory factors. Molecular and Cellular Biology, 12, 1031-1042.

Hua, X.-Y., E. Theodorsson-Norheim, E. Brodin, J.M. Lundberg, and T. Horkfelt (1985). Multiple tachykinins (neurokinin A, neuropeptide K and substance P) in capsaicin-sensitive sensory neurons in the guinea-pig. Regulatory Peptides, 13, 1-19.

Imagawa, M., R. Chiu, and M. Karin (1987). Transcription factor AP-2 mediates induction by two different signal-transduction pathways: protein kinase C and cAMP. Cell, 51, 251-260.

Ingles, C.J., J. Biggs, K.C. Wong, J.R. Weeks, and A.L. Greenleaf (1983). Identification of a structural gene for an RNA polymerase II polypeptide in *Drosophila melanogaster* and mammalian species. Proceedings of the National Academy of Sciences. USA., 80, 3396-3400.

Ingles, C.J., H.J. Himmelfrab, M. Shales, A.L. Greenleaf, and J.D. Friesen (1984). Identification, molecular cloning and mutagenisis of *Saccharomyces cerevisiae* RNA polymerase genes. Proceedings of the National Academy of Sciences. USA., 81, 2157-2161.

Ingles, C.J., M. Shales, W.D. Cress, S.J. Triezenberg, and J. Greenblatt (1991). Reduced binding of TFIID to transcriptionally comprimised mutants of VP16. Nature, 351, 588-590.

Ingraham, H.A., R. Chen, H.J. Mangalam, H.P. Elsholtz, S.E. Flynn, C.R. Lin, D.M. Simmons, L. Swanson, and M.G. Rosenfeld (1988). A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. Cell, 55, 519-529.

Ingraham, H.A., S.E. Flynn, J.W. Floss, V.R. Albert, M.S. Kaplioff, L. Wilson, and M.G. Rosenfeld (1990). The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1-Pit-1 interactions. Cell, 61, 1021-1033.

Inostroza, J., O. Flores, and D. Reinberg (1991). Factors involved in specific transcription by mammalian RNA polymerase II. Journal of Biological Chemistry, 266, 9304-9308.

Inostroza, J.A., F.H. Mermelstein, I. Ha, W.S. Lane, and D. Reinberg (1992). Dr1, a TATA-binding protein-associated phosphoprotein and inhibitor of class II gene transcription. Cell, 70, 477-489.

Ivashkiv, L.B., H. Liou, C.J. Kara, W.W. Lamph, I.M. Verma, and L.H. Glimcher (1990). mXBP/CRE-BP2 and c-Jun form a complex which binds to the cyclic AMP, but not to the 12-O-tetradecanoylphorbol-13-acetate, response element. Molecular and Cellular Biology, 10, 1609-1621.

Jaehning, J.A. (1991). Sigma factor relatives in eukaryotes. Science, 253, 859.

Janson, L., and U. Pettersson (1990). Cooperative interactions between transcription factors Sp1 and OTF-1. Proceedings of the National Academy of Sciences. USA., 87, 4732-4736.

Jantzen, H.-M., A. Admon, S.P. Bell, and R. Tjian (1990). Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. Nature, 344, 830-836.

Javahery, R., A. Khachi, K. Lo, B. Zenzic-Gregor, and S.T. Smale (1994). DNA sequence requirements for transcriptional initiator activity in mammalian cells. Molecular and Cellular Biology, 14, 116-127.

Jeang, K.T., R. Chun, N.H. Lin, A. Gatignal, C.G. Glabe, and H. Fan (1993). In vitro and in vivo binding of human immunodeficiency virus type 1 Tat protein and Sp1 transcription factor. Journal of Virology, 67, 6224-6233.

Jessel, T.M., L.L. Iversen, and A.C. Cuello (1978). Capsaicin-induced depletion of substance P from primary sensory neurons. Brain Research, 152, 183-188.

Johanson, T., U. Moens, T. Holm, A. Fjose, and S. Krauss (1993). Zebrafish *pou(c)*: a divergent POU family gene ubiquitously expressed during embryogenesis. Nucleic Acids Research, 21, 475-483.

Johnson, E.M., K.M. Rich, and H.K. Yip (1986). The role of NGF in sensory neurons. Trends in Neurosciences, 9, 33-37.

Johnson, P.F., and S.L. McKnight (1989). Eukaryotic transcriptional regulatory proteins. Annual Review of Biochemistry, 58, 799-839.

Johnson, J.E., S.J. Birren, and D.J. Anderson (1990). Two rat homologues of *drosophila achaete-scute* specifically expressed in neuronal precursors. Nature, 346, 858-861.

Johnson, J.E., S.J. Birren, T. Saito, and D.J. Anderson (1992). DNA binding and transcriptional regulatory activity of mammalian *achaete-scute* homologous (MASH) proteins revealed by interaction with a muscle-specific enhancer. Proceedings of the National Academy of Sciences. USA., 89, 3596-3600.

Johnston, B.H. (1988). The S1-sensitive form of d(C-T)*n*.d(A-G)*n:* chemical evidence for a three-stranded structure in plasmids. Science, 241, 1800-1804.

Jonakait, G.M., and S. Schotland (1990). Conditioned medium from activated splenocytes increases substance P in sympathetic ganglia. Journal of Neuroscience Research, 26, 24-30.

Jonakait, G.M., S. Schotland, and R.P. Hart (1990). Interleukin-1 specifically increases substance P in injured sympathetic ganglia. In: Neuropeptides and Immnunopeptides: messengers in the neuroimmune axis, 222-230.

Jones, K.A., and R. Tjian (1985). Sp1 binds to promoter sequences and activates herpes simplex virus 'immediate-early' gene transcription *in vitro*. Nature, 317, 179-182.

Jones, K.A., J. T. Kadonaga, P.A. Luciw, and R. Tjian (1986). Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. Science, 232, 755-759.

Jones, K.A., J.T. Kadonaga, P.J. Rosenfeld, T.J. Kelly, and R. Tjian (1987). A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell, 48, 79-89.

Jones, N. (1990). Transcriptional regulation by dimerisation: two sides to an incestious relationship. Cell, 61, 9-11.

Jordan, S.R., and C.O. Pabo (1988). Structure of the lambda complex at 2.5A resolution: details of the repressor-operator interactions. Science, 242, 893-899.

Kadonaga, J.T., K.A. Jones, and R. Tjian (1986). Promoter-specific activation of RNA polymerase II transcription by Sp1. Trends in Biochemistry, 11, 20-32.

Kadonaga, J.T., K.R. Carner, F.R. Masiarz, and R. Tjian (1987). Isolation of a cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell, 51, 1079-1090.

Kadonaga, J.T., A.J. Courey, J. Ladika, and R. Tjian (1988). In vitro analysis of Sp1 reveals distinct regions that modulate DNA binding and transcription. Science, 242, 1566-1570.

Kage, R., G.P. McGregor, L. Thim, and J.M. Conolon (1988). Neuropeptide-γ: a peptide isolated from rabbit intestine that is derived from γ-preprotachykinin. Journal of Neurochemistry, 50, 1412-1417.

Kageyama, R., Y. Sasai, and S. Nakanishi (1991). Molecular characterisation of transcription factors that bind to the cAMP responsive region of the substance P precursor gene. Journal of Biological Chemistry, 266, 15525-15531.

Kao, C.C., P.M. Lieberman, M.C. Schmidt, Q. Zhou, R. Pei, and A.J. Berk (1990). Cloning of a transcriptionally active human TATA binding factor. Science, 248, 1646-1650.

Kashanchi, F., G. Piras, M.F. Radonovich, J.F. Duvall, A. Fattaey, C.-M. Chiang, R.G. Roeder, and J.N. Brady (1994). Direct interaction of human TFIID with the HIV-1 transactivator Tat. Nature, 367, 295-299.

Kato, H., M. Horikoshi, and R.G. Roeder (1991). Repression of HIV-1 transcription by a cellular protein. Science, 251, 1476-1479.

Kato, G.J., W.M.F. Lee, L. Chen, and C.V. Dang (1992). Max: functional domains and interaction with c-Myc. Genes and Developmant, 6, 81-92.

Kawaguchi, Y., M. Hoshimaru, H. Nawa, and S. Nakanishi (1986). Sequence analysis of cloned cDNA for rat substance P precursor: existence of a third substance P precursor. Biochemical and Biophysical Research Communications, 139, 1040-1046.

Keen, P., A.B. Tullo, W.A. Blyth, and T.J. Hill (1982). Substance P in the mouse cornea: effects of chemical and surgical intervention. Neuroscience Letters, 29, 231-235.

Kefalas, P., F.C. Gray, and J. Allan (1988). Precise nucleosome positioning in the promoter of the chicken β^A-globin gene. Nucleic Acids Research, 16, 501-517.

Kelleher, R.J., S. Passmore, and A.D. Johnson (1989). Yeast repressor alpha 2 binds to its operator cooperatively with yeast protein Mcm1. Molecular and Cellular Biology, 9, 5228-5230.

Kelleher, R.J., P.M. Flanagan, D. I. Chasman, A.S. Ponticelli, K. Struhl, and R.D. Kornberg (1992). Yeast and human TFIID's are interchangable for the response to acidic transcriptional activators *in vitro*. Genes and Development, 6, 296-303.

Kemler, I., and W. Schaffner (1990). Octamer transcription factors and the cell-type specificty of immunoglobulin gene expression. FASEB Journal, 4, 1444-1449.

Kemp, L.M., and D.S. Latchman (1989). Regulated expression of herpes simplex virus immediate-early genes in neuroblastoma cells. Virology, 171, 607-610.

Kemp, L.M., C.L. Dent, and D.S. Latchman (1990). Octamer motif mediates transcriptional repression of HSV immediate-early genes and octamer-containing cellular promoters in neuronal cells. Neuron, 4, 215-222.

Kerppola, T.K., and T. Curran (1991). Transcription factors interactions: basics on leucine zippers. Current Opinions On Structural Biology.

Kessler, J.A., J.E. Adler, M.C. Bohn, and I.B. Black (1981). Substance P in principal sympathetic neurons: regulation by impulse activity. Science, 214, 335-336.

Killeen, M.T., and J.F. Greenblatt (1992). The general transcription factor RAP 30 binds to RNA polymerase II and prevents it from binding nonspecifically to DNA. Molecular and Cellular Biology, 12, 30-37.

Kinniburgh, A.J. (1989). A *cis*-acting transcription element of the c-*myc* gene can assume an H-DNA conformation. Nucleic Acids Research, 17, 7771-7778.

Kissinger, C.R., B.S. Liu, E. Martin-Blanco, T.B. Kornberg, and C.O. Pabo (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8A resolution: a framework for understanding homeodomain-DNA interactions. Cell, 63, 579-590.

Kitajima, S., Y. Tanaka, T. Kawaguchi, T. Nagaoka, S.M. Weissman, and Y. Yasukochi (1990). A heteromeric transcription factor required for mammalian RNA polymerase II. Nucleic Acids Research, 18, 4843-4849.

Klug, A., and D. Rhodes (1987). 'Zinc fingers': a novel protein motif for nucleic acid recognition. Trends in Biochemistry, 12, 464-469.

Ko, H.-S., P. Fast, W. McBride, and L.M. Staudt (1988). A human protein specific for the immunoglobulin octamer DNA motif contains a functional homeobox domain. Cell, 55, 135-144.

Kochoyan, M., T. Havel, D. Nguyen, C.E. Dahl, H.T Keutmann, and M. A. Weiss (1991). Alternating zinc fingers in the human male associated protein ZFY: 2D NMR

structure of an even finger and implications for "jumping-linker" DNA recognition. Biochemistry, 30, 3371-3386.

Kohwi, Y., and T. Kohwi-Shigematsu (1988). Magnesium ion-dependent triple-helix structure formed by homopurine-homopyrimidine sequences in supercoiled plasmid DNA. Proceedings of the National Academy of Sciences. USA., 85, 3781-3785.

Kohwi, Y. (1989). Cationic metal-specific structures adopted by the poly(dG) region and the direct repeats in the chicken adult β^A-globin gene promoter. Nucleic Acids Research, 17, 4493-4502.

Kohwi-Shigematsu, T., and Y. Kohwi (1985). Poly(dG)-poly(dG) sequences, under torsional stress, induce an altered DNA conformation upon neighboring DNA sequences. Cell, 43, 199-206.

Kolluri, R., and A.J. Kinniburgh (1991). Full length cDNA sequence encoding a nuclease-sensitive element DNA binding protein. Nucleic Acids Research, 19, 4771.

Kolluri, R.V., T.A. Torrey, and A.J. Kinniburgh (1992). A CT promoter element binding protein: definition of a double-strand and a novel single-strand DNA binding motif. Nucleic Acids Research, 20, 111-116.

Kotani, H., M. Hoshimaru, H. Nawa, and S. Nakanishi (1986). Structure and gene organization of bovine neuromedin K precursor. Proceedings of the National Academy of Sciences. USA., 83, 7074-7078.

Kovary, K., and R. Bravo (1991). Expression of different Jun and Fos proteins during the G_0-to-G_1 transition in mouse fibroblasts: in vitro and in vivo associations. Molecular and Cellular Biology, 11, 2451-2459.

Krause, J.E., J.M. Chirgwin, M.S. Carter, Z.S. Xu, and A.D. Hershey (1987). Three rat preprotachykinin mRNAs encode the neuropeptides substance P and neurokinin A. Proceedings of the National Academy of Sciences. USA., 84, 881-885.

Krause, J.E., J.D. Cremins, M.S. Carter, E.R. Brown, and M.R. MacDonald (1989). Solution hybridization-nuclease protection assays for sensitive detection of differentially spliced substance P- and neurokinin A-encoding messanger ribonucleic acids. Methods in Enzymology, 168, 634-652.

Kretzner, L., E.M. Blackwood, and R.N. Eisenman (1992). Myc and Max proteins posses distinct transcriptional activities. Nature, 359, 426-429.

Kruijer, W., D. Schubert, and I.M. Verma (1985). Induction of the proto-oncogene *fos* by nerve growth factor. Proceedings of the National Academy of Sciences. USA., 82, 7330-7334.

Kumar, V., S. Green, G. Stack, M. Berry, J.-R. Jin, and P. Chambon (1987). Functional domains of the human estrogen receptor. Cell, 51, 941-951.

Kumar, V., and P. Chambon (1988). The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer. Cell, 55, 145-156.

Kuraishi, Y., N. Hiroto, Y, Sato, Y. Hiro, M. Satoh, and H. Takagi (1985). Evidence that substance P and somatostatin transmit separate information related to pain in the spinal dorsal horn. Brain Research, 325, 294-298.

Kurkowski, R., W. Kummer, and C. Heym (1990). Substance P-immunoreactive nerve fibers in tracheobronchial lymph nodes of the guinea pig: origin, ultrastructure and coexistence with other peptides. Peptides, 11, 13-20.

Kutoh, E., P.E. Stromstedt, and L. Poellinger (1992). Functional interference between the ubiquitous and constitutive octamer transcription factor (OTF-1) and the glucocorticoid receptor by direct protein-protein interaction involving the homeo subdomain of OTF-1. Molecular and Cellular Biology, 12, 4960-4969.

Landschultz, W.H., P.F. Johnson, and S.L. McKnight (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding protein. Science, 240, 1759-1764.

Landschultz, W.A., P.F. Johnson, and S.L. McKnight (1989). The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science, 243, 1681-1688.

Lannigan, D.A., and A.C. Notides (1989). Estrogen receptor selectively binds the "coding strand" of an estrogen responsive element. Proceedings of the National Academy of Sciences. USA., 86, 863-867.

Larsen, A. and H. Weintraub (1982). An altered DNA conformation detected by S1 nuclease occurs at specific regions in active chick globin chromatin. Cell, 29, 609-622.

Lassar, A.B., R.L. Davis, W.E. Wright, T. Kadesch, C. Murre, A. Voronova, D. Baltimore, and H. Weintraub (1991). Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. Cell, 66, 305-315.

Laughon, A., and M.P. Scott (1984). Sequence of a Drosophila segmentation gene: protein structure homology with DNA-binding proteins. Nature, 320, 25-31.

LeBowitz, J.H., R.G. Clerc, M. Brenowitz, and P. Sharp (1989). The Oct-2 protein binds cooperatively to adjacent octamer sites. Genes and Development, 3, 1625-1638.

Lee, H.-L., and T.K. Archer (1994). Nucleosome-mediated disruption of transcription factor-chromatin initiation complexes at the mouse mammary tumour virus long terminal repeat in vivo. Molecular and Cellular Biology, 14, 32-41.

Lee, M.S., G.P. Gippert, K.V. Soman, D.A. Case, and P.E. Wright (1989). Three-dimensional solution structure of a single zinc finger DNA-binding domain. Science, 245, 635-637.

Leiberman, P.M., M.C. Schmidt, C.C. Kao, and A.J. Berk (1991). Two distinct domains in the yeast transcription factor IID and evidence for a TATA box-induced conformation change. Molecular and Cellular Biology, 11, 63-74.

Leiting, B., R. DeFrancesco, L. Tomei, R. Cortese, G. Otting, and K. Wuthrich (1993). The three-dimensional NMR-solution structure of the polypeptide fragment 195-286 of the LFB1/HNF1 transcription factor from liver comprises a non-classical homeodomain. The EMBO Journal, 12, 1797-1803.

Lembeck, F. (1953). Zur frage der zentralen ubertragung afferenter impulse. III. Das vorkommen und die bedeutung der substanz P in der dorsalen wurzelen des ruckenmarks. Naunyn-Schmiederbergs Archives of Experimental Pathology and Pharmacology.

Lembeck, F., J. Donnerer, and F.G. Colpaert (1981). Increase of substance P in primary afferent nerves during chronic pain. Neuropeptides, 1, 175-180.

Leuther, K.K., J.M. Salmeron, and S.A. Johnston (1993). Genetic evidence that an activation domain of GAL4 does not require acidity and may form a β sheet. Cell, 72, 575-585.

Levens, D., and P.M. Howley (1985). Novel method for identifying sequence-specific DNA-binding proteins. Molecular and Cellular Biology, 5, 2307-2315.

Lewis, C.D., S.P. Clark, G. Felsenfeld, and H. Gould (1988). An erythrocyte-specific protein that binds to the poly(dG) region of the chicken β-globin gene promoter. Genes and development, 2, 863-873.

Li, L., J.C. Chambard, M.Karin, and E.N. Olsen (1992). Fos and Jun repress transcriptional activation by myogenin and MyoD: the amino terminus of Jun can mediate repression. Genes and Development, 6, 676-689.

Li, H., and Y. Capetanaki (1993). Regulation of the mouse desmin gene: transactivation by MyoD, myogenin, MRF4 and Myf5. Nucleic Acids Research, 21, 335-343.

Lin, Y.-S., and M.R. Green (1991). Mechanism of action of an acidic transcriptional activator in vitro. Cell, 64, 971-981.

Lin, Y.-S., I. Ha, E. Maldonado, D. Reinberg, and M.R. Green (1991). Binding of general transcription factor TFIIB to an acidic activating region. Nature, 353, 569-571.

Lindefors, N., E. Brodin, and U. Ungerstedt (1989). Subchronic neuroleptic treatment decreases in vivo release of tachykinins in rat substantia nigra. European Journal of Pharmacology, 161, 95-98.

Lindefors, N. (1992). Amphetamine and haloperidol modulate preprotachykinin A mRNA expression in rat nucleus accumbens and caudate putamen. Molecular Brain Research, 13, 151-154.

Lindsay, R.M., and A.J. Harmar (1989). Nerve growth factor regulates expression of neuropeptide genes in adult sensory neurons. Nature, 337, 362-364.

Littlewood, T.D., B. Amati, H. Land, and G.I. Evan (1992). Max and c-Myc/Max DNA binding activities in cell extracts. Oncogene, 7, 1783-1792.

Lobo, S.M., M. Tanaka, M.L. Sullivan, and N. Hernandez (1992). A TBP complex essential for transcription from TATA-less but not TATA-containing RNA polymerase III promoters is part of the TFIIIB fraction. Cell, 71, 1029-1040.

Lonetto, M., M. Gribscov, and C.A. Gross (1992). The σ^{70} family: sequence conservation and evolutionary relationships. Journal of Bacteriology, 174, 3843-3849.

Lu, H., O. Flores, R. Weinmann, and D. Reinberg (1991). The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. Proceedings of the National Academy of Sciences. USA., 88, 10004-10008.

Lu, H., L. Zawel, L. Fisher, J.-M. Egly, and D. Reinberg (1992). Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. Nature, 358, 641-645.

Luisi, B.F., W.X. Xu, Z. Otwinowski, L.P. Freedman, K.R. Yamamoto, and P.B. Sigler (1991). Crystallographic analysis of the interaction of the glucacorticoid receptor with DNA. Nature, 352, 497-505.

Lundberg, J.M., T. Hokfelt, G. Nillson, L. Trenils, J. Rehfeld, R. Elde, and S. Said (1987). Peptide neurons in the vagus, splanchnic and sciatic nerves. Acta Physiologica Scandinavica, 104, 499-501.

Lyamichev, V.I., S.M. Mirkin, and M.D. Frank-Kamenetskii (1985). A pH-dependent structural transition in the homopurine-homoprymidine tract in superhelical DNA. Journal of Biomolecular Structural Dynamics, 2, 327-338.

Lyamichev, V.I., S.M. Mirkin, and M.D. Frank-Kamenetskii (1986). Structures of homopurine-homopyrimidine tract in superhelical DNA. Journal of Biomolecular Structural Dynamics, 3, 667-669.

Ma, J., and M. Ptashne (1987a). A new class of yeast transcriptional activators. Cell, 51, 113-119.

Ma, J., and M. Ptashne (1987b). Deletion analysis of GAL4 defines two transcriptional activating segments. Cell, 48, 847-853.

Ma, D., H. Watanabe, F. Mermelstein, A. Admon, K. Oguri, X. Sun, T. Wada, T. Imai, T. Shiroya, D. Reinberg, and H. Handa (1993). Isolation of a cDNA encoding the largest subunit of TFIIA reveals functions important for activated transcription. Genes and Development, 7, 2246-2257.

MacDonald, M.R., J. Takeda, C.M. Rice, and J.E. Krause (1989). Multiple tachykinins are produced and secreted upon post-translational processing of the three substance P precursor proteins, α-, β- and γ- preprotachykinin. Expression of the preprotachykinins in AtT-20 cells infected with vaccinia virus recombinants. Journal of Biological Chemistry, 264, 15578-15592.

Mack, D.H., J. Vartikar, J.M. Pipas, and L.A. Laimins (1993). Specific repression of TATA-mediated but not initiator mediated tanscription by wild-type p53. Nature, 363, 281-283.

Maekawa, T., H. Sakura, C. Kanei-Ishii, T. Sudo, T. Yoshimura, J.I. Fujisawa, M. Yoshida, and S. Ishii (1989). Leucine zipper structure of the protein CRE-BP1 binding to the cAMP response element in brain. The EMBO Journal, 8, 2023-2028.

Maki, Y., T.J. Bos, C. Davis, M. Starbuck, and P.K. Vogt (1987). Avian sarcoma virus 17 carries the *jun* oncogene. Proceedings of the National Academy of Sciences. USA., 84, 2848-2852.

Malik, S., K. Hisatake, H. Sumimoto, M. Horikoshi, and R.G. Roeder (1991). Sequence of general transcription factor TFIIB and relationships to other initiation factors. Proceedings of the National Academy of Sciences. USA., 88, 9553-9557.

Malthe-Sorensen, D., and O. Oktedalon (1982). Substance P-like immunoreactivity in the phrenic nerve. Neuroscience Letters, 28, 269-274.

Manley, J.L., A. Fire, A. Cano, P.A. Sharp, and M.L. Gefter (1980). DNA dependent transcription of adenovirus genes in a soluble whole cell extract. Proceedings of the National Academy of Sciences. USA., 77, 3855-3859.

Marcu, K.B., S.A. Bossone, and A.J. Patel (1992). Myc function and regulation. Annual Review of Biochemistry, 61, 809-860.

Mardon, G., and D.C. Page (1989). The sex-determining region of the mouse Y chromosone encodes a protein with a highly acidic domain and 13 Zinc fingers. Cell, 56, 765-770.

Masu, Y., K. Nakayama, H. Tamaki, Y. Harada, M. Kuno, and S. Nakanishi (1987). cDNA cloning of bovine substance K receptor through oocyte expression system. Nature, 329, 836-838.

Matsui, T., J. Segall, P.A. Weil, and R.G. Roeder (1980). Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. Journal of Biological Chemistry, 255, 11992-11996.

Matsui, M., M. Tokuhara, Y.N.N. Konuma, and R. Ishizaki (1990). Isolation of human fos-related genes and their expression during monocyte-macrophage differentiation. Oncogene, 5, 249-255.

Matthews, B.W., D.H. Ohlendorf, W.F. Anderson, and Y. Takeda (1982). Structure of the DNA-binding region of the *lac* repressor inferred from its homology with *cro* repressor. Proceedings of the National Academy of Sciences. USA., 79, 1428-1432.

Maue, R.A., S.D. Kraner, R.H. Goodman, and G. Mandel (1990). Neuron-specific expression of the rat brain type II sodium channel gene: a model for neuronal-specific gene regulation. Neuron, 4, 223-231.

Maurer, R.A., and Notides, A.C. (1987). Identification of an estrogen-responsive element from the 5'-flanking region of the rat prolactin gene. Molecular and Cellular Biology, 7, 4247-4254.

McAllister, J., S. Mendelson, and J.P. Quinn (1993). The preprotachykinin A promoter interacts with a sequence specific single stranded DNA binding protein. Biochemical Society Transactions, 21, 373S.

McCraken, S., and J.F. Greenblatt (1991). Related RNA polymerase-binding regions in human RAP 30/74 and *Escherichia coli* σ70. Science, 253, 900-902.

McGregor, G.P., R. Hartel, H.C. Fehmann, B. LankatButtgereit, B. Goke, R. Goke, and K. Voigt (1992). Characterisation of the expression and post-translational processing of the preprotachykinin-I gene and the regulated release of tachykinins by the RINm5F cell-line. FEBS Letters, 312, 187-191.

McKay, D.B., and T.A. Steitz (1981). Structure of the catabolite gene activator protein at 2.9A resolution suggests binding to left-handed B-DNA. Nature, 290, 744-749.

McKnight, S.L, and R. Kingsbury (1982). Transcriptional control signals of a eukaryotic protein encoding gene. Science, 217, 316-324.

Meisterernst, M., and R.G. Roeder (1991). Family of proteins that interact with TFIID and regulate promoter activity. Cell, 67, 557-567.

Meisterernst, M., A.L. Roy, H.M. Lieu, and R.G. Roeder (1991). Activation of class II gene transcription by regulatory factors is potentiated by a novel activity. Cell, 66, 981-993.

Merchant, J.L., B. Demediuk, and S.J. Brand (1991). A GC- rich element confers epidermal growth factor responsiveness to transcription from the gastrin promoter. Molecular and Cellular Biology, 11, 2686-2696.

Mermod, N., E.A. O'Neill, T.J. Kelley, and R. Tjian (1989). The proline-rich transcriptional activator of CTF/NF-1 is distinct from the replication and DNA binding domain. Cell, 58, 741-753.

Messier, H., H. Brickner, J. Gaikwad, and A. Fotedar (1993). A novel POU domain protein which binds to the T-cell receptor β enhancer. Molecular and Cellular Biology, 13, 5450-5460.

Milbrandt, J. (1986). Nerve growth factor rapidly induces c-fos mRNA in PC12 rat pheochromocytoma cells. Proceedings of the National Academy of Sciences. USA., 83, 4789-4793.

Milbrandt, J. (1987). A nerve growth factor-induced gene encodes a possible transcriptional regulatory factor. Science, 238, 797-799.

Milbrandt, J. (1988). Nerve growth factor induces a gene homologous to the gluacorticoid receptor gene. Neuron, 1, 183-188.

Miller, J., A.D. McLachlan, and A. Klug (1985). Repetitive Zn binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. The EMBO Journal., 4, 1609-1614.

Minamino, N., H. Masuda, K. Kangawa, and H. Matsuo (1984). Regional distribution of neuromedin K and neuromedin A in rat brain and spinal cord. Biochemical and Biophysical Research Communications, 114, 541-548.

Mitchell, P.J., C. Wang, and R. Tjian (1987). Positive and negative regulation of transcription in vitro: enhancer-binding protein AP-2 is inhibited by SV40 T antigen. Cell, 50, 847-861.

Mitchell, P.J., P.M. Timmons, J.M. Hebert, P.W.J. Rigby, and R. Tjian (1991). Transcription factor AP-2 is expressed in neuronal crest cell lineages during mouse embryogenesis. Genes and Development, 5, 105-119.

Mizrahi, J., S. Dion, P. D'Orleans-Juste, E. Escher, G. Drapeau, and D. Regoli (1985). Tachykinin receptors in smooth muscle: a study with agonists (substance P, neurokinin A) and antagonists. European Journal of Pharmacology, 118, 25-36.

Moncollin, V., L. Fischer, B. Cavillini, J.-M. Egly, and P. Chambon (1992). Class II (B) general transcription factor (TFIIB) that binds to the template-commited preinitiation complex is different from general transcription factor BTF3. Proceedings of the National Academy of Sciences. USA., 89, 397-401.

Morgan, J.I., and T. Curran (1991). Stimulus-transcription coupling in the nervous system: involvement of the inducible proto-oncogenes *fos* and *jun*. Annual Review of Neuroscience, 14, 421-451.

Moss, L.G., J.B. Moss, and W.J. Rutter (1988). Systematic binding analysis of the insulin gene transcription control region: insulin and immunoglobulin enhancers utilize similar transactivators. Molecular and Cellular Biology, 8, 2620-2627.

Morrison, C.F., J. McAllister, S.P. Dobson, P.K. Mulderry, and J.P. Quinn (1994). An activator element within the preprotachykinin-A promoter. Molecular and Cellular Neuroscience, in press.

Mueller, C.R., P. Maire, and U. Schibler (1990). DBP, a liver-enriched transcriptional activator, is expressed late in ontogeny and its tissue specificity is determined posttranscriptionally. Cell, 61, 279-291.

Muesing, M.A., D.H. Smith, and D.J. Capon (1987). Regulation of mRNA accumulation by a human immunodeficiency virus *trans*-activator protein. Cell, 48, 691-701.

Muhich, M.L., C.T. Iida, M. Horikoshi, R.G. Roeder, and C.S. Parker (1990). cDNA clone encoding Drosophila transcription factor TFIID. Proceedings of the National Academy of Sciences. USA., 87, 9148-9152.

Mukherjee, R., and P. Chambon (1990). A single-stranded DNA-binding protein promotes the binding of purified oestrogen receptor to its responsive element. Nucleic Acids Research, 18, 5713-5716.

Mulderry, P.K., K.E. Chapman, V. Lyons, and A.J. Harmar (1993). 5'-flanking sequences from the rat preprotachykinin gene direct high-level expression of a reporter gene in adult rat sensory neurons transfected in culture by microinjection. Molecular and Cellular-Neurosciences, 4, 164-172.

Muller-Immergluck, M.M., S. Ruppert, W. Schaffner, and P. Matthias (1988). A cloned transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. Nature, 336, 544-551.

Murre, C., P.S. McCaw, and D. Baltimore (1989a). A new DNA binding and dimerization motif in immunological enhancer binding, *daughterless*, *MyoD*, and *myc* proteins. Cell, 56, 777-783.

Murre, C., P.S. McCaw, H. Vaessin, M. Caudy, L.Y. Jan, Y. N. Jan, C.V. Cabrera, J.N. Buskin, S.D. Hauschka, A.B. Lassar, H. Weintraub, and D. Baltimore (1989b). Interaction between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell, 58, 537-544.

Nagai, K. (1990). Cryptic initiation sequence revealed. Nature, 343, 418.

Nagy, J.I., S.P. Hunt, L.L. Iverson, and P.C. Emson (1981). Biochemical and anatomical observations on the degeneration of peptide-containing primary afferent neurons after neonatal capsaicin. Neuroscience, 6, 1923-1934.

Nakabeppu, Y., K. Ryder, and D. Nathans (1988). DNA binding activities of three murine Jun proteins: stimulation by Fos. Cell, 55, 907-915.

Nakajima, N., M. Horikoshi, and R. Roeder (1988). Factors involved in specific transcription by mammalian RNA polymerase II, genetic specificty and TATA box-promoter interactions of TFIID. Molecular and Cellular Biology, 8, 4020-4040.

Nakatani, Y., M. Horikoshi, M. Brenner, T. Yamamoto, F. Besnard, R.G. Roeder, and E. Freese (1990). A downstream initiation element required for efficient TATA box binding and *in vitro* function of TFIID. Nature, 348, 86-88.

Nardelli, J., T. Gibson, C. Vesque, and P. Charnay (1991). Base sequence discrimination by zinc-finger DNA-binding domains. Nature, 349, 175-178.

Nawa, H., T. Hirose, H. Takashima, S. Inayama, and S. Nakanishi (1983). Nucleotide sequences of cloned cDNAs for two types of bovine substance P precursor. Nature, 306, 32-36.

Nawa, H., H. Kotani, and S. Nakanishi (1984). Tissue specific generation of two preprotachykinin mRNAs from one gene by alternative RNA splicing. Nature, 312, 729-734.

Nawa, H., and D.W.Y. Sah (1990). Different biological activities in conditioned media control the expression of a variety of neuropeptides in cultured sympathetic neurons. Neuron, 4, 279-287.

Nickol, J.M., and G. Felsenfeld (1983). DNA conformation at the 5' end of the chicken adult β-globin gene. Cell, 35, 467-477.

Nicosia, A., P. Monaci, L. Tomei, R. De Francesco, M. Nuzo, H. Stunnenberg, and R. Cortese (1990). A myosin-like dimerization helix and an extra-large homeodomain are essential elements of the tripartite DNA binding structure of LFB1. Cell, 61, 1225-1236.

Nilsson, J., A. Von Euler, and C.-J. Dalsgaard (1985). Stimulation of connective tissue cell growth by substance P and neurokinin A. Nature, 315, 61-63.

Nishina, H., H. Sato, T. Suzuki, M. Sato, and H. Iba (1990). Isolation and characterisation of *fra-2*, an additional member of the *fos* gene family. Proceedings of the National Academy of Sciences. USA., 87, 3619-3623.

Noguchi, K., Y. Monta, H. Kiyama, K. Ono, and M. Tohyama (1988). A noxious stimulus induces preprotachykinin-A gene expression in the rat dorsal root ganglion: a quantitative study using in situ hybridization histochemistry. Molecular Brain Research, 4, 31-35.

Noguchi, K., and M.A. Ruda (1992). Gene regulation in an ascending nociceptive pathway: inflammation induced increase in preprotachykinin mRNA in rat lamina 1 spinal projection neurons. Journal of Neuroscience, 12, 2563-2572.

Nonet, M., D. Sweetser, and R.A. Young (1987). Functional redundancy and structural polymorphism in the largest subunit of RNA polymerase II. Cell, 50, 909-915.

Ogata, K., H. Hojo, S. Aimoto, T. Nakai, H. Nakamura, A. Sarai, S. Ishii, and Y. Nishimura (1992). Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrphobic core. Proceedings of the National Academy of Sciences. USA., 89, 6428-6432.

Ogawa, T., I. Kanagawa, and S. Kimura (1985). Regional distribution of substance P, neurokinin α and neurokinin β in rat spinal cord, nerve roots and dorsal root ganglia, and the effects of dorsal root section or spinal cord transection. Brain Research, 359, 152-157.

O'Hare, P., and C.R. Goding (1988). Herpes simplex virus regulatory elements and the immunoglobulin octamer domain bind a common factor and are both targets for virion transactivation. Cell, 52, 435-445.

O'Hare, P., and G. Williams (1992). Structural studies of the acidic transactivation domain of the Vmw65 protein of herpes simplex virus using ^{3}H NMR. Biochemistry, 31, 4150-4156.

Ohkuma, Y., H. Sumimoto, M. Horikoshi, and R.G. Roeder (1990). Factors involved in specific transcription by mammalian RNA polymerase II: purification and characterisation of general transcription factor TFIIE. Proceedings of the National Academy of Sciences. USA., 87, 9163-9167.

Ohkuma, Y., H. Sumimoto, A. Hoffman, S. Shimasaki, M. Horikoshi, and R.G. Roeder (1991). Structrual motifs and potential σ homologies in the large subunit of human general transcription factor TFIIE. Nature, 354, 398-401.

Ohlendorf, D.H., W.F. Anderson, R.G. Fisher, Y. Takeda, and B.W. Mathews (1982). The molecular basis of DNA-protein recognition inferred from the structure of cro repressor. Nature, 298, 718-723.

Ohlendorf, D.H., W.F. Anderson, M. Lewis, C.O. Pabo, and B.W. Matthews (1983). Comparison of the structures of cro and λ repressor proteins from bacteriophage λ. Journal of Molecular Biology, 169, 757-769.

Okamoto, K., H. Okazawa, A. Okuda, M. Sakai, M. Muramatsu, and H. Hamada (1990). A novel octamer binding protein transcription factor differentially expressed in mouse embryonic cells. Cell, 60, 461-472.

Oliphant, A.R., C.J. Brandl, and K. Struhl (1989). Defining the sequence specificty of DNA-binding proteins by selective binding sites from random sequence oligonucleotides: analysis of yeast GCN4 protein. Molecular and Cellular Biology, 9, 2944-2949.

O'Neil, K.T., R.H. Hoess, and W.F. DeGrado (1990). Design of DNA binding peptides based on the leucine zipper motif. Science, 249, 774-778.

O'Shea, E.K., R. Rutkowski, and P.S. Kim (1989). Evidence that the leucine zipper is a coiled-coil. Science, 243, 538-542.

O'Shea, E.K., J.D. Klemm, P.S. Kim, and T. Alber (1991). X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science, 254, 539-544.

Otsuka, M., S. Konishi, and T. Takahashi (1972). The presence of motorneuron-depolarising peptide in bovine dorsal roots of spinal nerves. Proceedings of the Japenese Academy, 48, 342-346.

Otsuka, M., and S. Konishi (1976). Release of substance P-like immunoreactivity from isolated spinal cord of new born rats. Nature, 294, 83-84.

Otsuka, M., S. Konishi, M. Yanagisawa, A. Tsunoo, and H. Akagi (1982). Role of substance P as a sensory transmitter in the spinal cord and sympathetic ganglia. CIBA Foundation Symposium 91: Substance P in the nervous system (R.Porter and M.O'Connor, Eds.), Pitman, London, 13-30.

Otting, G., Y.Q. Qian, M. Billeter, M. Muller, M. Affolter, W.J. Ghering, and K. Wuthrich (1990). Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. The EMBO Journal, 9, 3085-3092.

Ouzounis, C., and C. Sander (1992). TFIIB, an evolutionary link between the transcriptional machineries of archaebacteria and eukaryotes. Cell, 71, 189-190.

Pabo, C.O., and M. Lewis (1982). The operator-binding domain of λ repressor: structure and DNA recognition. Nature, 298, 443-447.

Pan., W.T., Q. Lui, and C. Bancroft (1990). Identification of a growth hormone gene promoter repressor element and its cognate double- and single-stranded DNA-binding proteins. Journal of Biological Chemistry, 265, 7022-7028.

Paonessa, G., F. Gounari, R. Frank, and R. Cortese (1988). Purification of an NF-1 like DNA binding protein from rat liver and cloning of the corresponding cDNA. The EMBO Journal, 7, 3115-3123.

Papka, R.E., J.B. Furness, N.G. Della, R. Murphy, and M. Costa (1984). Time course of effect of capsaicin on ultra-structure and histochemistry of substance P-immunoreactive nerves associated with the cardiovascular system of the guinea-pig. Neuroscience, 4, 1277-1292.

Park, K., and K.H. Kim (1993). The site of cAMP action in the insulin induction of gene expression of acetyl-CoA carboxylase is AP-2. Journal of Biological Chemistry, 268, 17811-17819.

Park, E.A., A.L. Gurney, S.E. Nizielski, P. Hakimi, Z. Cao, A. Moorman, and R.W. Hanson (1993). Relative roles of CCAATT/enhancer-binding protein β and cAMP regulatory element-binding protein in controlling transcription of the gene for phosphoenolpyruvate carboxykinase (GTP). Journal of Biological Chemistry, 268, 613-619.

Parraga, G., S.J. Horvath, A. Eisen, W.E. Taylor, L. Hood, E.T. Young, and R.E. Klevit (1988). Zinc-dependent structure of a single-finger domain of yeast ADR1. Science, 241, 1489-1492.

Patel, L., C. Abate, and T. Curran (1990). Altered protein conformation on DNA binding by Fos and Jun. Nature, 347, 572-574.

Pavletich, N.P., and C.O. Pabo (1991). Zinc finger-DNA recognition: crystal structure of a Zif 268-DNA complex at 2A. Science, 252, 809-817.

Pavletich, N.P., and C.O. Pabo (1993). Crystal structure of a five-finger GL1-DNA complex: new perspectives on zinc fingers. Science, 261, 1701-1707.

Payne, J.M., P.J. Laybourn, and M. Dahmus (1989). The transition of RNA polymerase II from initiation to elongation is associated with phosphorylation of the carboxy-terminal domain of subunit IIa. Journal of Biological Chemistry, 264, 19621-19629.

Pendergast, G.C., D. Lawe, and E.B. Ziff (1991). Association of Myn, the murine homologue of Max, with c-Myc stimulates methylation-sensitive DNA binding and Ras cotransformation. Cell, 65, 395-407.

Pernow, B. (1985). Role of tachykinins in neurogenic inflammation. Journal of Immunology, 135, 812s.

Peterson, M.G., N. Tanese, B.F. Pugh, and R. Tjian (1990). Functional domains and upstream activator properties of cloned human TATA binding protein. Science, 248, 1625-1630.

Peterson, M.G., J. Inostroza, M.E. Maxon, O. Flores, A. Admon, D. Reinberg, and R. Tjian (1991). Structure and functional properties of human general transcription factor IIE. Nature, 354, 369-373.

Peterson, S.R., A. Dvir, C.W. Anderson, and W.S. Dynan (1992). DNA binding provides a signal for phosphorylation of the RNA polymerase II heptapeptide repeats. Genes and Development, 6, 426-438.

Phillipe, J., D.J. Drucker, W. Knepel, L. Jepeal, Z. Misulovin, and J.F. Habener (1988). Alpha-cell-specific expression of the glucagon gene is conferred to the glucagon promoter element by the interactions of DNA-binding proteins. Molecular and Cellular Biology, 8, 4877-4888.

Poellinger, L., and R.G. Roeder (1989). Octamer transcription factors 1 and 2 each bind to two different functional elements in the immunoglobulin heavy-chain promoter. Molecular and Cellular Biology, 9, 747-756.

Poellinger, L., B.K. Yoza, and R.G. Roeder (1989). Functional cooperativity between protein molecules bound to two distinct sequence elements of the immunoglobulin heavy-chain promoter. Nature, 337, 573-576.

Poon, D., S. Schroeder, C.K. Wang, T. Yamamota, M. Horikoshi, R.G. Roeder, and P.A. Weil (1991). The conserved carboxy-terminal domain of *Saccharomyces cerevisae* TFIID is sufficient to support normal cell growth. Molecular and Cellular Biology, 11, 4809-4820.

Popper, P., C.R. Mantyh, S.R. Vigna, J.E. Magio, and P.W. Mantyh (1988). The localization of sensory nerve fibers and receptor binding sites for sensory neuropeptides in canine mesenteric lymph nodes. Peptides, 9, 257-267.

Postel, E.H., S.E. Mango, and S.J. Flint (1989). A nuclease-hypersensitive element of the human *c-myc* promoter interacts with a transcription initiation factor. Molecular and Cellular Biology, 9, 5123-5133.

Preston, C.M. (1979). Control of herpes simplex virus type 1 mRNA synthesis in cells infected with wild-type virus or the temperature sensitive mutant tsK. Journal of Virology, 29, 275-284.

Price, J. (1985). An immunohistochemical and quantitative examination of dorsal root ganglion neuronal populations. Journal of Neuroscience, 5, 2051-2059.

Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. Nature, 322, 697-701.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. Nature, 335, 683-689.

Pugh, B.F., and R. Tjian (1990). Mechanism of transcriptional activation by Sp1: evidence for coactivators. Cell, 61, 1187-1197.

Pugh, B.F., and R. Tjian (1991). Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes and Development, 5, 1935-1945.

Pugh, B.F., and R. Tjian (1992). Diverse transcriptional functions of the multisubunit eukaryotic TFIID complex. Journal of Biological Chemistry, 267, 679-682.

Qian, Y.Q., M. Billeter, G. Otting, M. Muller, W.J. Gehring, and K. Wuthrich (1989). The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. Cell, 59, 573-580.

Quinn, J.P., N. Holbrook, and D. Levens (1987). Binding of a cellular protein to the Gibbon Ape leukemia virus enhancer. Molecular and Cellular Biology, 7, 2735-2744.

Quinn, J.P. (1991). Variation in the composition of the AP1 complex in PC12 cells following induction by NGF and TPA. Molecular and Cellular Neurosciences, 2, 253-258.

Quinn, J.P. (1992). Multiple protein binding sites within the rat preprotachykinin promoter: demonstration of a site with neuronal specificity that is 3' of the major transcriptional start site. Molecular and Cellular Neurosciences, 3, 11-16.

Quinn, J.P., and J.McAllister (1993). The preprotachykinin promoter interacts with a sequence specific single stranded DNA binding protein. Nucleic Acids Research, 21, 1637-1641.

Randic, M., and V. Miletic (1977). Effect of substance P on cat dorsal horn neurones activated by noxious stimuli. Brain Research, 128, 164-169.

Ranish, J.A., and S. Hahn (1991). The yeast general transcription factor TFIIA is composed of two polypeptide subunits. Journal of Biological Chemistry, 266, 19320-19327.

Ranish, J.A., W.S. Lane, and S. Hahn (1992). Isolation of two genes that encode subunits of the yeast transcription factor TFIIA. Science, 255, 1127-1129.

Rao, M.S., S. Tyrrell, S.C. Landis, and P.H. Patterson (1992). Effects of ciliary neurotrophic factor (CTNF) and depolarization on neuropeptide expression in cultured sympathetic neurons. Developmental Biology, 150, 281-293.

Rauscher III, F.J., L.C. Sambucetti, T. Curran, R.J. Distel, and B.M. Spiegelman (1988). A common DNA binding site for Fos protein complexes and transcription factor AP1. Cell, 52, 471-480.

Reddy, P., and S. Hahn (1991). Dominant negative mutations in yeast TFIID define a bipartite DNA-binding region. Cell, 65, 349-357.

Reinberg, D., M. Horikoshi, and R.G. Roeder (1987). Factors involved in specific transcription by mammalian RNA polymerase II. Journal of Biological Chemistry, 262, 3322-3330.

Renkawitz, R., G. Schultz, D. Von Der Ahe, and M. Beato (1984). Sequences in the promoter of the chicken lysozyme gene required for steroid regulation and receptor binding. Cell, 37, 503-510.

Revoltella, R., L. Bertolini. M. Pediconi, and E. Vigneti (1974). Specific binding of nerve growth factor (NGF) by murine C1300 neuroblastoma cells. The Journal of Experimental Medicine, 140, 437-451.

Rey-Campos, J., T. Chouard, M. Yaniv, and S. Cereghini (1991). vHNF1 is a homeoprotein that activates transcription and forms heterodimers with HNF1. The EMBO Journal, 10, 1445-1457.

Riegel, A.T., J. Remenick, R.G. Wolford, D.S. Berard, and G.L. Hager (1990). A novel transcriptional activator (PO-B) binds between the TATA box and the cap site of the pro-opiomelanocortin gene. Nucleic Acids Research, 18, 4513-4521.

Rio, D.C., and R. Tjian (1983). SV40 T antigen binding site mutations that affect autoregulation. Cell, 32, 1227-1240.

Ritter, J.K., C.J. Schmidt, J.W. Gibb, and G.R. Hanson (1984). Increases of substance P-like immunoreactivity within striatal-nigral structures after subacute metamphatamine treatment. Journal of Pharmacological and Experimental Therapeutics, 229, 487-492.

Roach, A., J.E. Adler, and I.B. Black (1987). Depolarizing influences regulate preprotachykinin mRNA in sympathetic neurons. Proceedings of the National Academy of Sciences. USA., 84, 5078-5081.

Roberts, S.G.E., I. Ha, E. Maldonado, D. Reinberg, and M.R. Green (1993). Interaction between an acidic activator and transcription factor TFIIB is required for transcriptional activation. Nature, 363, 741-744.

Roesler, W.J., G.R. Vandenbark, and R.W. Hanson (1988). Cyclic AMP and the induction of eukaryotic gene transcription. Journal of Biological Chemistry, 263, 9063-9066.

Roizman, B., and A.E. Sears (1986). An enquiry into the mechanism of herpes simplex virus latency. Annual Review Microbiology, 41, 543-571.

Roman, C., J.S. Platro, J. Shuman, and K. Calame (1990). A ubiquitously expressed immunoglobulin enhancer binding protein that is similar to C/EBP and heterodimerises with C/EBP. Genes and Development, 4, 1404-1415.

Ron, D, and J.F. Habener (1992). CHOP, a novel developmentally regulated nuclear protein that dimerizes with transcription factors C/EBP and LAP and functions as a dominant-negative regulator of gene transcription. Genes and Development, 6, 439-453.

Rosen, C.A., J.G. Sodroski, and W.A. Haseltine (1985). The location of cis-acting regulatory sequences in the human T-cell lymphocyte virus type III (HTLV-III/LAV) long terminal repeat. Cell, 41, 813-823.

Rosenberg, U.B., C. Schroder, A. Preiss, A. Keinlin, S. Cote, I. Riede, and H. Jackle (1986). Structural homology of the product of the *Drosophila Kruppel* gene with *Xenopus* transcription factor IIIA. Nature, 319, 336-339.

Rosenfeld, M.G. (1991). POU-domain transcription factors: pou-er-ful developmental regulators. Genes and Development, 5, 897-907.

Roy, A.L., M. Meisterernst, P. Pognonec, and R.G. Roeder (1991). Cooperative interactions of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. Nature, 354, 245-248.

Ruppert, S., E.H. Wang, and R. Tjian (1993). Cloning and expression of human $TAF_{II}250$: a TBP-associated factor implicated in cell-cycle regulation Nature, 362, 175-179.

Ryder, K., L.F. Lau, and D. Nathans (1988). A gene activated by growth factors is related to the oncogene v-*jun*. Proceedings of the National Academy of Sciences. USA., 85, 1487-1491.

Ryder, K., A. Lanahan, E. Perez-Albuerne, and D. Nathans (1989). *Jun*-D: a third member of the *Jun* gene family. Proceedings of the National Academy of Sciences. USA., 86, 1500-1503.

Samuels, M., and P.A. Sharp (1986). Purification and characterisation of a specific RNA polymerase II transcription factor. Journal of Biological Chemistry, 261, 2003-2013.

Santhanam, U., A. Ray, and P.B. Sehgal (1991). Repression of the interleukin 6 gene promoter by p53 and the retinoblastoma susceptibility gene product. Proceedings of the National Academy of Sciences. USA., 88, 7605-7609.

Santoro, C., N. Mermod, P.C. Andrews, and R. Tjian (1988). A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. Nature, 334, 218-224.

Santoro, I.M., T.-M. Yi, and K. Walsh (1991). Identification of single-stranded-DNA-binding proteins that interact with muscle gene elements. Molecular and Cellular Biology, 11, 1944-1953.

Saria, A., E. Theodorsson-Norheim, R. Gamse, and J.M. Lundberg (1984). Release of substance P- and substance K-like immunoreactivity from the isolated perfused guinea-pig ileum. European Journal of Pharmacology, 106, 207-208.

Saudek, V., A. Pastore, M.A. Castiglione-Morelli, R. Frank, H. Gausepohl, and T. Gibson (1991). The solution structure of a leucine-zipper motif peptide. Protein Engineering, 4, 519-529.

Sauer, R.T., R.R. Yocum, R.F. Doolittle, M. Lewis, and C.O. Pabo (1982). Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. Nature, 298, 447-451.

Sauer, R.T., D.L. Smith, and A. Johnson (1988). Flexibility of the yeast $\alpha 2$ repressor enables it to occupy the ends of its operator, leaving the center free. Genes and Development, 2, 807-816.

Sauer, R.T., S.R. Jordan, and C.O. Pabo (1990). Lambda repressor: a model system for understanding protein-DNA interactions and protein stability. Advances in Protein Chemistry, 40, 1-61.

Sawadogo, M., and R.G. Roeder (1985a). Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative *in vitro* assay. Proceedings of the National Academy of Sciences. USA., 82, 4394-4398.

Sawadogo, M., and R.G. Roeder (1985b). Interaction of a gene specific transcription factor with the adenovirus major late promoter of the TATA box region. Cell, 43, 165-175.

Sawadogo, M., and A. Sentenac (1990). RNA polymerase B II and general transcription factors. Annual Review of Biochemistry, 59, 711-754.

Sayre, M.H., H. Tschochner, and R.D. Kornberg (1992). Reconstitution of transcription with five purified initiation factors and RNA polymerase II from *saccharomyces cerevisiae*. Journal of Biological Chemistry, 267, 23376-23382.

Schaffner, W. (1989). How do different transcription factors binding the same DNA sequence sort out their jobs?. Trends in Genetics, 5, 37-39.

Scheidereit, C., J.A. Cromlish, T. Gerster, K. Kawakami, C.-G. Balmaceda, R.A. Currie, and R.G. Roeder (1988). A human lymphoid-specific transcription factor that activates immunoglobulin genes in a homeobox protein. Nature, 336, 552-557.

Schmidt, M.C., C.C. Kao, R. Pei, and A.J. Burk (1989a). Yeast TATA-box transcription factor gene. Proceedings of the National Academy of Sciences. USA., 86, 7785-7789.

Schmidt, M.C., Q. Zhou, and A.J. Berk (1989b). Sp1 activates transcription without enhancing DNA-binding activity of the TATA box factor. Molecular and Cellular Biology, 9, 3299-32307.

Schneuwly, S., A. Kuroiwa, P. Baumgartner, and W.J. Gehring (1986). Structural organization and sequence of the homeotic gene *Antennapedia* of *Drosophila* melanogaster. The EMBO Journal, 5, 733-739.

Scholer, H.R., A.K. Hatzopoulos, R. Balling, N. Suzuki, and P. Gruss (1989). A family of octamer-specific proteins present during mouse embryogenesis: evidence for germ-line specific expression of an Oct factor. The EMBO Journal, 8, 2543-2550.

Scholer, H.R., S. Ruppert, N. Suzuki, K. Chowdhury, and P. Gruss (1990). New type of POU domain in germ-line specific protein Oct-4. Nature, 334, 435-439.

Schreiber, E., P. Matthias, M.N. Muller, and W. Schaffner (1988). Identification of a novel lymphoid specific octamer binding protein (OTF-2B) by proteolytic clipping bandshift assay (PCBA). The EMBO Journal, 7, 4211-4229.

Schreiber, E., K. Harshman, I. Kemler, U. Malipiero, W. Schaffner, and A. Fontana (1990). Astrocytes and glioblastoma cells express novel octamer-DNA binding proteins distinct from the ubiquitous Oct-1 and B cell type Oct-2 proteins. Nucleic Acids Research, 18, 5495-5503.

Schulz, S.C., G.C. Shields, and T.A. Steitz (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90^{o}. Science, 253, 1001-1007.

Schwabe, J., D. Neuhaus, and D. Rhodes (1990). Solution structure of the DNA-binding domain of the oestrogen receptor. Nature, 348, 458-461.

Scott, M.P., J.W. Tamkun, and G.W. Hartzell (1989). The structure and function of the homeodomain. Biochim. Acta, 989, 25-48.

Sekiguchi, T., T. Miyata, and T. Nishimoto (1988). Molecular cloning of the cDNA of human X chromosonal gene (CCG1) which complements the temperature-sensitive G_1 mutants, tsBN462 and ts13, of the BHK cell line. The EMBO Journal, 7, 1683-1687.

Sekiguchi, T.,Y. Nohiro, Y. Nakamura, N. Hisamoto, and T. Nishimoto (1991). The human CCG1 gene, essential for progression of the G1 phase, encodes a 210-kilodalton nuclear DNA-binding protein. Molecular and Cellular Biology, 11, 3317-3325.

Serfling, E., M. Jasin, and W. Schaffner (1985). Enhancers and eukaryotic gene transcription. Trends in Genetics, 1, 224-230.

Seto, E., Y. Shi, and T. Shenk (1991). YY1 is an initiator sequence-binding protein that directs and activates transcription *in vitro*. Nature, 345, 241-245.

Seto, E., A. Usheva, G.P. Zambetti, J. Momad, N. Horikoshi, R. Weinmann, A. J. Levine, and T. Shenk (1992). Wild-type p53 binds to the TATA-binding protein and represses transcription. Proceedings of the National Academy of Sciences. USA., 89, 12028-12032.

Shadiack, A.M., R.P. Hart, C.D. Carlson, and G.M. Jonakait (1993). Interleukin-1 induces substance P in sympathetic ganglia through the induction of leukemia inhibitory factor (LIF). Journal of Neuroscience, 13, 2601-2609.

Sharp, P.A. (1992). TATA-binding protein is a classless factor. Cell, 68, 819-821.

Sheng, M., M.A. Thomson, and M.E. Greenberg (1991). CREB: a Ca^{2+} regulated transcripton factor phosphorylated by calmodulin-dependent kinases. Science, 252, 1427-1430.

Shepherd, J.C.W., W. McGinnis, A.E. Carrasco, E.M. De Robertis, and W.J. Gehring (1984). Fly and frog homeodomains show homologies with yeast mating type regulatory proteins. Nature, 310, 70-71.

Shi, Y., E. Seto, L.-S. Chang, and T. Shenk (1991). Transcriptional repression by YY1, a human GLI-kruppel-related protein, and relief of repression by adenovirus E1A protein. Cell, 67, 377-388.

Shigemoto, R., Y. Yokota, K. Tsuchida, and S. Nakanishi (1990). Cloning and expresion of a rat neuromedin K receptor cDNA. Journal of Biological Chemistry, 265, 623-628.

Siebenlist, U., L. Hennighausen, J. Battey, and P. Leder (1984). Chromatin structure and protein binding in the putative regulatory region of the *c-myc* gene in Burkitt lymphoma. Cell, 37, 381-391.

Sigler, P.B. (1988). Acid blobs and negative noodles. Nature, 333, 210-212.

Simon, A.M., and S.J. Burden (1993). An E Box mediates activation and repression of the acetylcholine receptor δ-subunit gene during myogenisis. Molecular and Cellular Biology, 13, 5133-5140.

Singer, V.L., C.R. Wobbe, and K. Struhl (1990). A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. Genes and Development, 4, 636-645.

Singh, H., R. Sen, D. Baltimore, and P.A. Sharp (1986). A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature, 319, 154-158.

Smale, S.T., and D. Baltimore (1989). The 'initiator' as a transcription control element. Cell, 57, 103-113.

Smale, S.T., M.C. Schmidt, A.J. Berk, and D. Baltimore (1990). Transcription activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. Proceedings of the National Academy of Sciences. USA., 87, 4509-4513.

Smith, G.D., J.R. Seckl, and A.J. Harmar (1993). Distribution of neuropeptides in dorsal root ganglia of the rat; substance P, somatostatin, and calcitonin gene-related peptide. Neuroscience Letters, 153, 5-8.

Sodroski, J.G., C.A. Rosen, W.C. Goh, and W.A. Haseltine (1985). A transcriptional activator protein encoded by the x-*lor* region of the human T-cell leukemia virus. Science, 228, 1430-1434.

Sopta, M., Z.F. Burton, and J. Greenblatt (1989). Structure and associated DNA-helicase activity of a general transcription factor that binds to RNA polymerase II. Nature, 341, 410-414.

Spencer, C.A., and M. Groudine (1990). Transcription elongation and eukaryotic gene regulation. Advances in Cancer Research, 56, 1-48.

Stark, H.C., O. Weinberger, and J. Weinberger (1992). Common double- and single-stranded DNA binding factor for a sterol regulatory element. Proceedings of the National Academy of Sciences. USA., 89, 2180-2184.

Staudt, L.M., R.G. Clerc, H. Singh, J.H. LeBowitz, P.A. Sharp, and D. Baltimore (1988). Cloning of a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif. Science, 241, 577-580.

Steitz, T.A., D.H. Ohlendorf, D.B. McKay, W.F. Anderson, and B.W. Matthews (1982). Structural similarities in the DNA-binding domains of catabolite gene activator and *cro* repressor proteins. Proceedings of the National Academy of Sciences. USA., 79, 3097-3100.

Stevens, J.G., E.K. Wagner, G.B. Devi-Rao, M.L. Cook, and L.T. Feldman (1987). RNA complementary to a herpes virus gene mRNA is prominent in latently infected neurons. Science, 235, 1056-1058.

Stoykova, A.S., S. Sterrer, J.R. Erselius, A.K. Hatzopoulos, and P. Gruss (1992). Mini-Oct and Oct-2c: two novel, functionally diverse murine Oct-2 gene products are differentially expressed in the CNS. Neuron, 8, 541-558.

Stringer, K.F., C.J. Ingles, and L. Greenblatt (1990). Direct and selective binding of an acidic transcriptional activation domain to the TATA box factor TFIID. Nature, 345, 783-786.

Struhl, K. (1987). The DNA-binding domains of the jun oncoprotein and the yeast GCN4 transcriptional activator protein are functionally homologous. Cell, 50, 841-846.

Struhl, K. (1989). Molecular mechanisms of transcriptional regulation in yeast. Annual Review of Biochemistry, 58, 1051-1077.

Sturm, R.A., and W. Herr (1988). The POU-domain is a bipartite DNA-binding structure. Nature, 336, 601-604.

Sturm, R.A., G. Das, and W. Herr (1988). The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. Genes and Development, 2, 1582-1599.

Su, W., S. Jackson, R. Tjian, and H. Echols (1991). DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. Genes and Development, 5, 820-826.

Suckow, M., B. von Wilken-Bergmann, and B. Muller-Hill (1993). Identification of three residues in the basic regions of the bZIP proteins GCN4, C/EBP and TAF-1 that are involved in specific DNA binding. The EMBO Journal, 12, 1193-1200.

Sukhatme, V.P., X. Cao, L.C. Chang, C. Tsai-Morris, D. Stamenkovich, P.C.P. Ferreira, D.R. Cohen, S.A. Edwards, T.B. Shows, T. Curran, M.M. LeBeau, and E.D. Adamson (1988). A Zn finger-encoding gene coregulated with c-*fos* during growth and differentiation, and after cell depolarization. Cell, 53, 37-43.

Sumimoto, H., Y. Ohkuma, T. Yamamoto, M. Horikoshi, and R.G. Roeder (1990). Interactions between RNA polymerase II, factors, and template leading to accurate transcription. Proceedings of the National Academy of Sciences. USA., 87, 9158-9162.

Suminoto, H., Y. Ohkuma, E. Sinn, H. Kato, S. Shimasaki, M. Horikoshi, and R.G. Roeder (1991). Conserved sequence motifs in the small subunit of human transcription factor TFIIE. Nature, 354, 401-404.

Sun, X.-H., and D. Baltimore (1991). An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. Cell, 64, 459-470.

Sun, X.H., N.G. Copeland, N.A. Jenkins, and D. Baltinore (1991). Id proteins Id1 and Id2 selectively inhibit DNA binding by one class of helix-loop-helix proteins. Molecular and Cellular Biology, 11, 5603-5611.

Suzuki, N., H. Rhodewohld, T. Neuman, P. Gruss, and H.R. Scholer (1990). Oct-6: a POU-domain transcription factor expressed in embryonal and stem cells and in the developing brain. EMBO Journal, 9, 3723-3732.

Svaren, J., and R. Chalkley (1990). The structure and assembly of active chromatin. Trends in Genetics, 6, 52-56.

Takahashi, T., and M. Otsuka (1975). Regional distribution of substance P in the spinal cord and the nerve roots of the cat and the effect of dorsal root section. Brain Research, 87, 1-11.

Takimoto, M., T. Tamonaga, M. Matunis, M. Avigan, H. Krutzsch, G. Dreyfuss, and D. Levens (1993). Specific binding of heterogeneous ribonucleoprotein particle protein K to the human c-myc promoter, in vitro. Journal of Biological Chemistry, 263, 18249-18258.

Talanian, R.V., C.J. McKnight, and P.S. Kim (1990). Sequence-specific DNA binding by a short peptide dimer. Science, 249, 769-771.

Tamura, T., K. Sumita, K. Fujino, A. Aoyama, M. Horikoshi, A. Hoffmann, R.G. Roeder, M. Muramatsu, and K. Mikoshiba (1991). Striking homology of the variable N-terminal as well as the conserved core domains of the mouse and human TATA-factors (TFIID). Nucleic Acids Research, 19, 3861-3865.

Tanese, N., B.F. Pugh, and R. Tjian (1991). Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. Genes and Development, 5, 2122-2124.

Tatemoto, K., J.M. Lundberg, H. Jornvall, and V. Mutt (1985). Neuropeptide K: isolation, structure and biological activities of a novel brain tachykinin. Biochemical and Biophysical Research Communications, 128, 947-953.

Theil, T., S. McLean-Hunter, M. Zornig, and T. Moroy (1993). Mouse BRN-3 family of POU transcription factors: a new aminoterminal domain is crucial for the oncogenic activity of BRN-3A. Nucleic Acids Research, 21, 5921-5929.

Thompson, N.E., T.H. Steinberg, D.B. Aronson, R.R. Burgess (1989). Inhibition of in vivo and in vitro transcription by monoclonal antibodies prepared against wheat germ RNA polymerase II that reacts with the heptapeptide repeat of eukaryotic RNA polymerase II. Journal of Biological Chemistry, 264, 11511-11520.

Timmers, H.T., and P.A. Sharp (1991). The mammalian TFIID protein is present in two functionally distinct complexes. Genes and Development, 5, 1946-1956.

Too, H.-P., J.L. Cordova, and J.E. Maggio (1989). Heterogeneity of tachykinin-like immunoreactive peptides in rat spinal cord and dorsal root ganglia. Peptides, 10, 25-30.

Traurig, H., R.E. Papka, A. Saria, and F. Lembeck (1984). Substance P immunoreactivity in the rat mammary nipple and the effect of capsaicin treatment on lactation. Nauyn-Schmiedeberg's Archives of Pharmacology, 328, 1-8.

Travis, A., A. Amsterdam, C. Berlanger, and R. Grosschedl (1991). LEF-1, a gene encoding a lymphoid-specific protein, with an HMG domain, regulates T-cell receptor α enhancer function. Genes and Development, 5, 880-894.

Treacy, M.N., X. He, and M.G. Rosenfeld (1991). I-POU: a POU-domain protein that inhibits neuron-specific gene activation. Nature, 350, 577-584.

Triesman, R. (1985). Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5' element and c-fos 3' sequences. Cell, 42, 889-902.

Triesman, R. (1986). Identification of a protein-binding site that mediates transcriptional response of the c-*fos* gene to serum factors. Cell, 46, 567-574.

Tsai, S.Y., M.J. Tsai, and B.W. O'Malley (1989). Cooperative binding of steroid hormone receptors contributes to transcriptional synergism at target enhancer elements. Cell, 57, 443-448.

Tschochner, H., M.H. Sayre, P.M. Flanagan, W.J. Feaver, and R.D. Kornberg (1992). Yeast RNA polymerase II initiation factor e: isolation and identification as the functional counterpart of human transcription factor IIB. Proceedings of the National Academy of Sciences. USA., 89, 11292-11296.

Tsuboi, A., K. Conger, K.P. Garret, R.C. Conaway, J.W. Conaway, and N. Arai (1992). RNA polymerase II initiation factor α from rat liver is almost identical to human TFIIB. Nucleic Acids Research, 20, 3250.

Tuchscherer, M.M., and V.S. Seybold (1985). Immunohistochemical studies of substance P, cholecystokinin octapeptide and somatostatin in dorsal root gangia of the rat. Neuroscience, 14, 593-605.

Tverberg, L.A., and A.F. Russo (1993). Regulation of the calcitonin/calcitonin gene-related peptide by cell-specific synergy between helix-loop-helix and octamer-binding transcription factors. Journal of Biological Chemistry, 268, 15965-15973.

Usheva A., E. Maldonado, A. Goldring, H. Lu, C. Houbavi, D. Reinberg, and Y. Aloni (1992). Specific interaction between the nonphosphorylated form of RNA polymerase II and the TATA-binding protein. Cell, 69, 871-881.

Usuda, Y., A. Kubota, A.J. Berk, and H. Handa (1991). Affinity purification of transcription factor IIA from HeLa cell nuclear extracts. The EMBO Journal, 10, 2305-2310.

Vallejo, M., D. Ron, C.P. Miller, and J.F. Habener (1993). C/ATF, a member of the activating transcription factor family of DNA-binding proteins, dimerizes with CAAT/enhancer-binding proteins and directs their binding to cAMP response elements. Proceedings of the National Academy of Sciences. USA., 90, 4679-4683.

Van Dyke, M.W., R.G. Roeder, and M. Sawadogo (1988). Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. Science, 241, 1335-1338.

Van Dyke, M.W., and M. Sawadogo (1990). DNA-binding and transcriptional properties of human transcription factor TFIID after mild proteolysis. Molecular and Cellular Biology, 10, 3415-3420.

Van Hoy, M., K.K. Leuther, T. Kodadek, and S.A. Johnson (1993). The acidic activation domains of the GCN4 and GAL4 proteins are not α helical but form β sheets. Cell, 72, 587-594.

Verrijzer, C.P., A.J. Kal, and P.C. van der Vliet (1990). The Oct-1 homeo domain contacts only part of the octamer sequence and full Oct-1 DNA-binding activity requires the POU-specific domain. Genes and Development, 4, 1964-1974.

Verrijzer, C.P., M.A. Alkema, W.W. van Weperen, H.C. Van Leeuwen, M.J.J. Strating, and P.C. van der Vilet (1992). The DNA binding specificity of the bipartite POU domain and its subdomains. The EMBO Journal, 11, 4993-5003.

Vinson, C.R., P.B. Sigler, and S.L. McKnight (1989). A scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science, 246, 911-916.

Vinson, C.R., and K.C. Garcia (1992). Molecular model for DNA recognition by the family of basic-helix-loop-helix-zipper proteins. N. Biol., 4, 396-403.

Voronova, A., and D. Baltimore (1990). Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. Proceedings of the National Academy of Sciences. USA., 87, 4722-4726.

Voss, J.W., L. Wilson, and M.G. Rosenfeld (1991). POU-domain proteins Pit-1 and Oct-1 interact to form a heteromeric complex and cooperate to induce regulation of the prolactin promoter. Genes and Development, 5, 1309-1320.

Wagner, S., and M.R. Green (1993). HTLV-1 Tax protein stimulation of DNA binding of bZIP proteins by enhancing dimerization. Science, 262, 395-399.

Walker, J.M., K. Gooderham, J.R.B. Hastings, E. Mayes, and E.W. Johns (1980). The primary structures of non-histone chromosonal proteins HMG1 and 2. FEBS Letters, 122, 264-270.

Wampler, S.L., and J.T. Kadonaga (1992). Functional analysis of *Drosophila* transcription factor IIB. Genes and Development, 6, 1542-1552.

Wang, W., M. Carey, and J.D. Gralla (1992). Polymerase II promoter activation; closed complex formation and ATP-driven start site opening. Science, 255, 450-453.

Watson, R.J., and J.B. Clements (1980). A herpes simplex virus type 1 function continuously required for early and late viral RNA synthesis. Nature, 285, 329-330.

Weber, I.T., D.B. McKay, and T.A. Steitz (1982). Two helix DNA binding motifs of CAP found in *lac* and *gal* repressor. Nucleic Acids Research, 10, 5085-5102.

Weihe, E., S. Muller, T. Fink, and H.J. Zentel (1989). Tachykinins, calcitonin gene-related peptide and neuropeptide Y in nerves of the mammalian thymus: interactions with mast cells in autonomic and sensory neuroimmunomodulation? Neuroscience Letters, 100, 77-82.

Weil, P.A., D.S. Luse, J. Segall, and R.G. Roeder (1979). Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II. Cell, 18, 469-484.

Weintraub, H., R. Davis, S. Tapscott, M. Thayer, M. Krause, R. Benezra, K. Blackwell, D. Turner, R. Rupp, S. Hollenberg, Y. Zhuang, and A. Lassar (1991). The MyoD gene family: nodal point during specification of the muscle-cell lineage. Science, 251, 761-766.

Weiss, M.A., T. Ellenberger, C.R. Wobbe, J.P. Lee, S.C. Harrison, and K. Struhl (1990). Folding transition in the DNA binding domain of GCN4 on specific binding to DNA. Nature, 347, 575-578.

Whelan, J., S.R. Cordle, E. Henderson, P.A. Weil, and R. Stein (1990). Identification of a pancreatic β-cell insulin gene transcription factor that binds to and appears to activate cell-type-specific expression: its possible relationship to other cellular factors that bind to a common insulin gene sequence. Molecular and Cellular Biology, 10, 1564-1572.

White, J.H., C. Brou, J. Wu, N. Burton, J.-M. Egly, and P. Chambon (1991). Evidence for a factor required for transcriptional stimulation by the chimeric acidic activator GAL-VP16 in HeLa cell extracts. Proceedings of the National Academy of Sciences. USA., 88, 7674-7678.

White, J., C. Brou, J. Wu, Y. Lutz, V. Moncollin, and P. Chambon (1992). The acidic transcriptional activator GAL-VP16 acts on preformed template-committed complexes. The EMBO Journal, 11, 2229-2240.

Wilcox, C.L., and E.M. Johnson (1987). Nerve growth factor deprivation results in the reactivation of latent herpes simplex virus in vitro. Journal of Virology, 61, 2311-2315.

Wilcox, C.L., and E.M. Johnson (1988). Characterization of nerve growth factor-dependent herpes simplex virus latency in neurons in vitro. Journal of Virology, 62, 393-399.

Wiley, S.R., R.J. Kraus, and J. Mertz (1992). Functional binding of the "TATA" box binding component of transcription factor TFIID to the -30 region of TATA-less promoters. Proceedings of the National Academy of Sciences. USA., 89, 5814-5818.

Wiley, S.R., R.J. Kraus, F. Zuo, E.E. Murray, and J.E. Mertz (1993). SV40 early-to-late switch involves titration of cellular transcriptional repressors. Genes and Development, 7, 2206-2219.

Wilkison, W.O., H.Y. Min, K.P. Claffey, B.L. Satterberg, and B.M. Spiegelman (1990). Control of the adipsin gene in adipocyte differentiation. Journal of Biological Chemistry, 265, 477-482.

Williams, T., A. Admon, B. Luscher, and R. Tjian (1988). Cloning and expression of AP-2, a cell-type-specific transcription factor that activates inducible enhancer elements. Genes and Development, 2, 1557-1569.

Williams, S.C., C.A. Cantwell, and P.F. Johnson (1991). A family of C/EBP-related proteins capable of forming covalently linked leucine zipper dimers in vitro. Genes and Development, 5, 1553-1567.

Williams, T., and R. Tjian (1991). Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. Genes and Development, 5, 670-682.

Wilson, D., G. Sheng, T. Lecuit, N. Dostatni, and C. Desplan (1993). Cooperative dimerization of paired class homeo domains on DNA. Genes and Development, 7, 2120-2134.

Wobble, C.R., and K. Struhl (1990). Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription *in vitro*. Molecular and Cellular Biology, 10, 3859-3867.

Wolberger, C., A.K. Vershon, B. Liu, A.D. Johnson, and C.O. Pabo (1991). Crystal structure of a Matα2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. Cell, 67, 517-528.

Womack, M.D., A.B. MacDermot, and T.M. Jessel (1988). Sensory transmitters regulate intracellular calcium in dorsal horn neurons. Nature, 334, 351-353.

Wood, J.N., S.J .Bevan, P.R. Coote, P.M. Dunn, A. Harmar, P. Hogan, D.S. Latchman, C. Morrison, and G. Rougon (1990). Novel cell lines display properties of nociceptive sensory neurons. Proceedings of the Royal Society of London, Series B: Biological Sciences (London), 241, 187-194.

Wood, J.N., K.A. Lillycrop, C.L. Dent, N.N. Ninkina, M.M. Beech, J.J. Willoughby, J. Winter, and D.S. Latchman (1992). Regulation of expression of the neuronal POU protein Oct-2 by nerve growth factor. Journal of Biological Chemistry, 267, 17787-17791.

Wright, W.E., D.A. Sassoon, and V.K. Linn (1989). Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. Cell, 56, 607-617.

Wright, W.E., M. Binder, and W. Funk (1991). Cyclic amplification and selection of targets (CASTing) for the myogenic consensus binding site. Molecular and Cellular Biology, 11, 4104-4110.

Xiang, M., L. Zhou, Y.-W., Peng, R.L. Eddy, T.B. Shows, and J. Nathans (1993). Brn-3: a POU domain gene expressed in a subset of retinal ganglion cells. Neuron, 11, 689-701.

Xue, D., Y. Tu, and M. Chalfie (1993). Cooperative interactions between the *Caenorhabditis elegans* homeoproteins Unc-86 and Mec-3. Science, 261, 1324-1328.

Yaksh, T.L., T.M. Jessel, R. Gamse, A.W. Mudge, and S.E. Leeman (1980). Intrethecal morphine inhibits substance P release from mammalian spinal cord in vitro. Nature, 280, 155-157.

Yamamota, K.K., G.A. Gonzalez, W.H. Biggs, and M.R. Montmony (1988). Phosphorylation-induced binding and transcriptional efficacy of nuclear factor CREB. Nature, 334, 494-498.

Yamamota, T., M. Horikoshi, J. Wang, S. Hasegawa, P.A. Weil, and R.G. Roeder (1992). A bipartite DNA binding domain composed of direct repeats in the TATA box binding factor TFIID. Proceedings of the National Academy of Sciences. USA., 89, 2844-2848.

Yokota, Y., Y. Sasai, K. Tanaka, T. Fujiwara, K. Tsuchida, R. Shigemoto, A. Kakizuka, H. Ohkubo, and S. Nakanishi (1989). Molecular characterization of a functional cDNA for rat substance P receptor. Journal of Biological Chemistry, 264, 17649-17652.

Yokoyama, C., X. Wang, M. R. Briggs, A. Admon, J. Wu, X. Hua, J.L. Goldstein, and M.S. Brown (1993). SREBP-1, a basic-helix-loop-helix zipper protein that controls transcription of the low density lipoprotein receptor gene. Cell, 75, 187-197.

Yoon, S.O., and D.M. Chikaraishi (1992). Tissue-specific transcription of the rat tyrosine hydroxylase gene requires synergy between an AP1 motif and an overlapping E box-containing dyad. Neuron, 9, 55-67.

Yoshida, T., Y. Shindo, K. Ohta, and H. Iba (1991). Transcription of fra-2 mRNA and phosphorylation of Fra-2 protein are stimulated by serum. Biochemical and Biophysical Research Communications, 174, 934-939.

Young, R.A. (1991). RNA polymerase II. Annual Review of Biochemistry, 60, 689-715.

Zambetti, G.P., J. Bargonetti, K. Walker, C. Prives, and A.J. Levine (1992). Wild-type p53 mediates positive regulation of gene expression through a specific DNA sequence element. Genes and Development, 6, 1143-1152.

Zehring, W.A., J.M. Lee, J.R. Weeks, R.S. Jokerst, and A.L. Greenleaf (1988). The C-terminal repeat domain of RNA polymerase II largest subunit is essential *in vivo* but is not required for accurate transcription *in vitro*. Proceedings of the National Academy of Sciences. USA., 85, 3698-3702.

Zeiglgansberger, W., and T.R. Tolle (1993). The pharmacology of pain signalling. Current Opinions in Neurobiology, 3, 611-618.

Zerial, M., L. Toschi, R.P. Ryseck, M. Schuermann, R. Miller, and R. Bravo (1989). The product of a novel growth factor activated gene, FosB, interacts with JUN proteins enhancing their DNA binding activity. The EMBO Journal, 8, 805-813.

Zervos, A.S., J. Gyuris, and R. Brent (1993). Mxi1, a protein that specifically interacts with Max to bind Myc-Max recognition sites. Cell, 72, 223-232.

Zhang, Y., J. Babin, A.L. Feldhaus, H. Singh, P.A. Sharp, and M. Bina (1991). HTF4: a new human helix-loop-helix protein. Nucleic Acids Research, 19, 4555.

Zhou, Q., M.C. Schmidt, and A.J. Berk (1991). Requirement for acid amino acids residues immediately N-terminal to the conserved domain of saccharomyces cerevisiae TFIID. The EMBO Journal, 10, 1843-1852.

Zhou, Q., P.M. Lieberman, T.J. Boyer, and A.J. Berk (1992). Holo-TFIID supports transcriptional stimulation by diverse activators and from a TATA-less promoter. Genes and Development, 6, 1964-1974.

Zhou, Q., T.G. Boyer, and A.J. Berk (1993). Factors (TAFs) required for activated transcription interact with TATA box-binding protein conserved domain. Genes and Development, 7, 180-187.

Molecular Brain Research, 15 (1992) 174–178

BRESM 80144

An NGF-inducible octamer binding protein activity in a C1300 neuroblastoma cell line

S.C. Mendelson, C. Morrison and J.P. Quinn

MRC Brain Metabolism Unit, Royal Edinburgh Hospital, Edinburgh (UK)

(Accepted 2 June 1992)

Key words: Nerve growth factor; Octamer binding protein

Nerve growth factor (NGF) induces a protein in the C1300 mouse neuroblastoma cell line which recognises the octamer DNA consensus sequence 'ATGCAAAT'. This protein is absent, or only minimally present, in C1300 cells prior to induction with NGF. This induced octamer binding protein is detectable by gel retardation analysis within 3 h of NGF treatment and increases progressively with 24 h and 72 h of exposure to NGF.

The octamer binding motif is a DNA consensus sequence of ATGCAAAT, which can activate transcription of both cell-specific and ubiquitously expressed genes[1,5,15,27,29]. In addition to being an activator of transcription it can be found within elements which repress transcription[17,33]. The ability of this sequence motif to elicit such a range of transcriptional responses is, in part, due to various degenerate DNA motifs being recognised by the octamer binding protein family, and also the large number of different octamer binding proteins and their own tissue-specific expression[2,19,26,28,34]. These octamer binding proteins are a subset of a larger family of regulatory molecules which share a protein domain referred to as the POU-domain[9,10,13,25,31]. The POU domain is responsible for binding the DNA and contains two regions, an N-terminal POU-specific domain and a C-terminal homeodomain. These 'POU' proteins have been shown in the developing and adult brain to have a restricted pattern of expression[12]. In the brain, or neuronal derived cells, at least five different octamer binding proteins exist[26,32]. The ubiquitous octamer binding protein Oct 1 has been demonstrated to be a target for phosphorylation[24], and Pit 1 the Pou domain containing transcription factor has its DNA binding affinity modulated by phosphorylation[14]. In light of the number of octamer binding proteins in the brain, and that they can be modulated by post-translational modification, we postulated that octamer binding proteins in neuronal cells may be regulated by nerve growth factor.

We have therefore analysed the neuroblastoma cell line C1300 for changes in octamer binding proteins in the presence or absence of nerve growth factor (NGF). The C1300 cell line does contain a receptor for NGF[11], and although addition of NGF to these cells does not differentiate the cells morphologically as with PC12 cells[8], changes can be observed in several parameters[3,6,11,23]. Using gel retardation, we demonstrate the presence of an NGF-inducible octamer binding protein in C1300 cells.

The DNA consensus sequence for the octamer binding protein from the immunoglobulin enhancer[1,29] was used to identify specific octamer binding protein complexes in the C1300 cell line by gel retardation analysis. Three complexes were observed in C1300 cells whereas an extra complex was observed when the cells were exposed to NGF. As judged by competition with a heterologous oligonucleotide the faster migrating complexes were non-specific binding complexes (indicated by N.S. in Fig. 1). These non-specific complexes have previously been demonstrated to bind DNA oligonucleotide probes, some of which contain consensus oc-

Correspondence: J.P. Quinn, MRC Brain Metabolism Unit, Royal Edinburgh Hospital, Morningside Park, Edinburgh EH10 5HF, UK.

tamer binding sequences[22,33]. The slower migrating complex in C1300 cells is indicated to be the ubiquitous octamer binding protein Oct 1 because its migration is in the same position as the Oct 1 complex from HeLa cells (Fig. 2B). Following exposure of the cells to NGF, a second specific complex, migrating slightly above the observed non-specific complexes, was observed (C1300 OCT, Fig. 1). This NGF-induced complex was more easily seen as the amount of non-specific oligonucleotide competitor was increased, removing the non-specific complexes (Fig. 1, lanes 6–10).

The specificity of the induced complex was demonstrated by competition with homologous and heterologous oligonucleotides (Fig. 2A,B). As no competition of the specific complexes was observed except where homologous double-stranded oligonucleotide was added, this confirmed that we have identified an NGF-inducible octamer binding protein. Although in Fig. 1 the intensity of the Oct 1 complex also increases when the cells are exposed to NGF, this effect was not reproducible and as observed in Fig. 2 we routinely saw no difference in Oct 1 binding activity.

Previous reports of a repressor protein present in C1300 cells which inhibited the activity of octamer consensus sequence containing promoters indicated that C1300 cells constitutively expressed two specific octamer binding complexes[16]. Therefore, we obtained C1300 cells from a different source and repeated our study, to demonstrate that our finding of an NGF-inducible octamer binding protein was not unique to our isolate of C1300 cells. Our results were found to be reproducible: only Oct 1 was constitutively present in C1300 cells and the second octamer binding protein was absent, or only minimally present, prior to exposure of the cells to NGF. This source of cells was used in Fig. 3, where a time course of induction by NGF of this second octamer binding protein is shown. This NGF-inducible octamer binding protein was first detected by gel retardation as a faint complex at 3 h post-exposure to NGF (data not shown), and was clearly observed at 48 and 72 h post-induction. The increase in the C1300 NGF inducible octamer binding activity is such that the intensity of this complex obscures the Oct 1 complexes at both 24 and 72 h post-exposure to NGF. The specificity of the complexes were demonstrated by competition with heterologous competitor (NS) or homologous competitor (WT). In initial experiments to determine whether this induced octamer binding had an effect on gene transcription, we used the HSV IE 4/5 promoter (which contains an octamer consensus sequence) to express the chloramphenicol acetyltransferase protein (CAT) (a gift from R. Everett). In agreement with previous reports of HSV immediate early genes being repressed in C1300 cells[16] we found no significant CAT activity in these cells above background of cells transfected with pGEM vector which contains no CAT gene (data not shown). Exposure of these cells to NGF had no effect on the CAT activity. Therefore, we are unable to correlate appearance of the NGF inducible octamer binding protein with repression of this promoter, and our results might indicate repression of activity in C1300 cells by a different mechanism.

We have demonstrated that NGF induces an octamer binding protein activity distinct from that of the Oct 1 protein in the mouse C1300 neuroblastoma cell line. The migration of this NGF-inducible octamer in

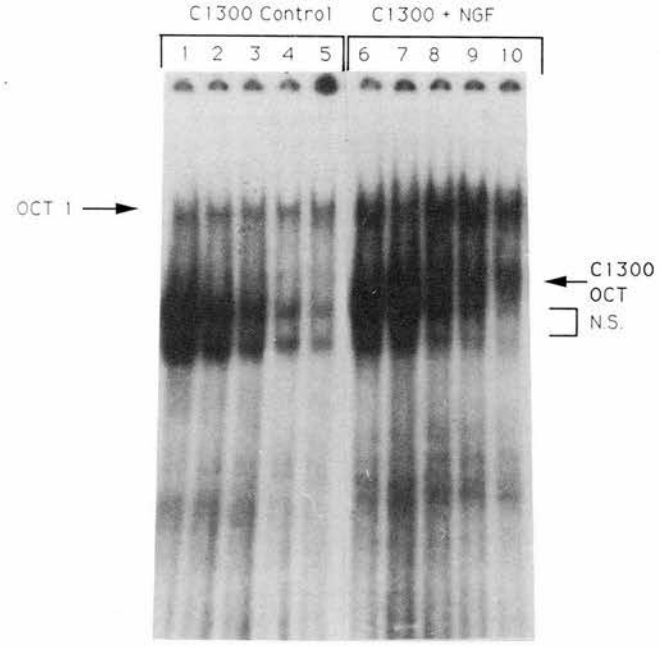


Fig. 1. C1300 cells contain an NGF-inducible octamer binding complex. Whole cell extracts were prepared from C1300 cells as previously described[18]. The protein extract was separated from the cell debris by centrifugation. Whole cell extracts contained approximately 20 $\mu g/\mu l$ protein. Gel retardation analysis was performed as described previously[7,27,30]. Briefly, each reaction contained 0.1–1 ng of ^{32}P 3′ end-labelled double-stranded oligonucleotide probe from the well-characterised immunoglobulin octamer binding element (5′AGATCTAGCATGCAAATCATTGT)[1,29]. Cell extracts (20 μg) were incubated with 1 μg of poly d[I-C] prior to addition of probe and competitor oligonucleotides. Extracts from C1300 cells and C1300 cells exposed to NGF (Collaborative Research) for 24 h, at a final concentration of 100 ng/ml, were subjected to gel retardation analysis. Lanes 1–5 contain C1300 cell extract and lanes 6–10 contain C1300 cell extract where cells had been exposed to NGF for 24 h. The position of the ubiquitous octamer complex, Oct 1, is indicated with an arrow. Exposure of cells to NGF gives rise to a new complex, C1300 Oct, which is indicated with an arrow, migrating above non-specific complexes (N.S.). Increasing concentration of non-specific oligonucleotide (5′ATCCCTTTAAATTTGCGAGCT) N.S. (10, 50, 100 and 150 ng) was added in lanes 2–5 and lanes 7–10, to remove non-specific complexes. This heterologous oligonucleotide sequestered DNA binding proteins which preferentially bind to the termini of DNA fragments [21].

C1300 cells would be similar to that obtained with the same DNA probe using purified Oct 2 from our previous studies[33]. However, several neuronal octamer binding proteins in brain nuclear extract, in particular, N-Oct α and β[26] and an octamer binding protein from malignant melanoma cells[4] all migrate in a similar position to Oct 2. We cannot determine the particular octamer binding protein from its migration in gel retardation analysis. Therefore, the nature of the C1300 NGF-induced protein remains to be characterised.

In a previous study of octamer binding proteins present in C1300 cells, evidence was presented for two distinct octamer binding complexes, one of which was presumably Oct 1[16]. It is not possible to directly compare our results with this study for a number of reasons: (i) it is possible that a number of strains of C1300 cells exist[6]; (ii) it has been demonstrated at least in PC12 cells, that the transcription factor complement of the cell can vary with growth in culture[20].

The NGF-induced octamer binding protein in C1300 cells is regulated temporally differently from the effect of NGF on other transcription factors, which have been found to be NGF-inducible, and termed immediate early proteins. These immediate early proteins, such as c-fos and c-jun, are turned on rapidly and transiently in response to NGF in PC12 cells[8,20]. The C1300-induced octamer binding protein appears to be present and increased long after the initial stimulus (Fig. 3), so it may fall into the class of proteins termed 'delayed early'. This protein would have the temporal characteristics of a protein which was maintaining a phenotype in response to the NGF stimulus. The fact that we demonstrate this in C1300 cells, which do not differentiate morphologically in response to NGF[3,6,23],

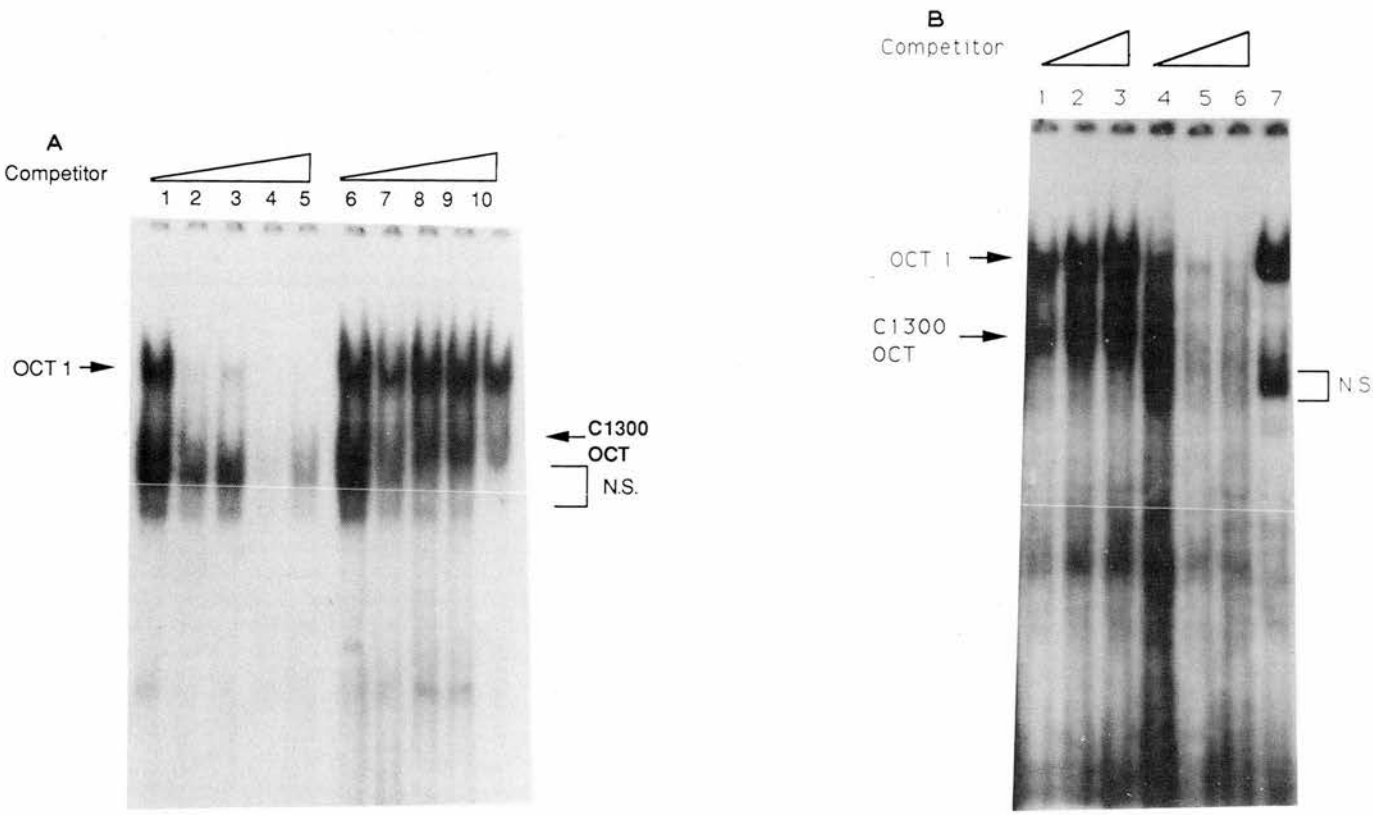


Fig. 2. A: the NGF-induced complex in C1300 cells is a specific octamer binding protein. Extracts from C1300 cells exposed to NGF for 24 h were subjected to gel retardation analysis. Non-specific oligonucleotide (50 ng) was pre-incubated with poly d[I-C] in the cell extract to remove non-specific complexes. The position of the Oct 1 and C1300 NGF-induced Oct shifts are indicated with arrows. Competitor was added to demonstrate the specificity of the NGF-induced complex. Lanes 1 and 6 contain no competitor. Lanes 2–5 contain 10, 25, 50 and 100 ng of homologous competitor and lanes 7–10 contain 10, 25, 50 and 100 ng of heterologous competitor, respectively (the same N.S. oligonucleotide as used in the preincubation). B: the NGF-induced complex in C1300 cells is a specific octamer binding protein. Extracts from C1300 cells exposed to NGF for 24 h were subjected to gel retardation analysis and the mobility of the retarded complexes compared with those obtained with HeLa cells which constitutively express Oct 1. Non-specific oligonucleotide (50 ng) was pre-incubated with poly d[I-C] in the cell extract to remove non-specific complexes. The position of the Oct 1 and C1300 NGF-induced Oct shifts are indicated with arrows. Competitor was added to demonstrate the specificity of the NGF-induced complex. In lanes 1, 2 and 3, 10, 25 and 50 ng of additional heterologous competitor was added. In lanes 4, 5 and 6, 10, 25 and 50 ng of homologous specific competitor was added, respectively. Lane 7 shows the ubiquitous Oct 1 shift in HeLa control cells.

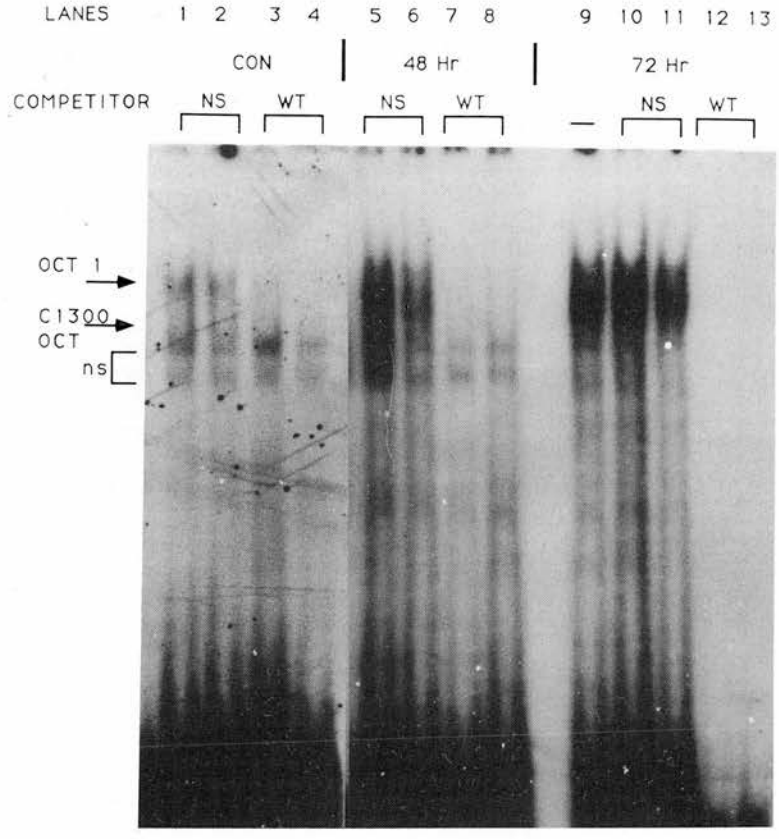


Fig. 3. The NGF-induced octamer complex is present at 48 and 72 h post-induction. A second source of C1300 cells (J. Morrow, University of Glasgow) were grown in DMEM with 10% heat-inactivated fetal calf serum. NGF (Collaborative Research) was used at a final concentration of 100 ng/ml. Extracts from C1300 cells and C1300 cells exposed to NGF for 48 and 72 h were subjected to gel retardation analysis. Lanes 1–4 contain C1300 cell extract. Lanes 5–8 contain C1300 cell extract where the cells had been exposed to NGF for 48 h and lanes 9–13 contain C1300 cell extract where the cells had been exposed to NGF for 72 h. The position of the Oct 1 and C1300 NGF-induced Oct complexes are indicated with arrows. Competitor oligonucleotides were added to demonstrate the specificity of the NGF-induced complex. In lanes 1 and 2, 5 and 6, 10 and 11, 25 and 50 ng of heterologous competitor was added, respectively. In lanes 3 and 4, 7 and 8, 12 and 13, 25 and 50 ng of homologous competitor was added, respectively.

may allow us to correlate the appearance of this octamer binding protein with gene regulation, rather than with cell differentiation. This aspect of C1300 cell regulation may make them a better model for study of NGF regulation of transcription rather than PC12 cells.

We wish to thank Drs. V.J. Bubb, A.J. Harmar, J. Paterson, and K. Reid for critical reading of the manuscript, and Norma Brearley for her careful preparation of the manuscript. S. Mendelson is the recipient of an MRC Student Fellowship.

1 Augereau, P. and Chambon, P., The mouse immunoglobulin heavy-chain enhancer: effect on transcription in vitro and binding of proteins present in HeLa and lymphoid B cell extracts, *EMBO J.*, 5 (1986) 1791–1797.
2 Baumrucker, T., Strum, R. and Herr, W., OPB100 binds remarkably degenerate octamer motifs through specific interactions with flanking sequences, *Gene Dev.*, 2 (1988) 1400–1413.
3 Brodeur, G.M. and Goldstein, M.N., Histochemical demonstration of an increase in acetylcholinesterase in established lines of human and mouse neuroblastomas by nerve growth factor, *Cytobios*, 16 (1976) 133–138.
4 Cox, P.M., Temperley, S.M., Kumar, H. and Goding, C.R., A distinct octamer-binding protein present in malignant melanoma cells, *Nucl. Acid Res.*, 16 (1988) 11047–11056.
5 Davidson, I., Fromental, C., Augereau, P., Wildeman, A., Zenke, M. and Chambon, P., Cell-type specific protein binding to the enhancer of simian virus 40 in nuclear extracts, *Nature*, 323 (1986) 544–548.
6 Diamond, L., Revoltella, R. and Bertolini, L., Differences between murine C1300 neuroblastoma clones detected by rosette formation with nerve growth factor-coated sheep red blood cells, *Brain Res.*, 118 (1976) 453–459.
7 Fried, M. and Crothers, D.M., Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis, *Nucl. Acids Res.*, 9 (1981) 6505–6525.
8 Fujita, K., Lazarovici, P. and Gurroff, G., Regulation of the differentiation of PC12 pheochromocytoma cells, *Exp. Health Perspect.*, 8 (1989) 127–142.
9 Garcia-Blanco, M.A., Clerc, R.G. and Sharp, P.A., The DNA-binding homeodomain of the Oct-2 protein, *Genes Dev.*, 3 (1989) 739–745.
10 Gehring, W.J., Müller, M., Affolter, M., Percival-Smith, A., Billeter, M., Quian, Y.Q., Otting, G. and Wüthrich, The structure of the homoeodomain and its functional implications, *TIG*, 6 (1990) 323–329.
11 Greene, L.A. and Shooter, E.M., The nerve growth factor: biochemistry, synthesis, and mechanism of action, *Ann. Rev. Neurosci.*, 3 (1980) 353–402.
12 He, X., Treacy, M.N., Simmons, D.M., Ingraham, H.A., Swanson, L.W. and Rosenfeld, M.G., Expression of a large family of POU-domain regulatory genes in mammalian brain development, *Nature*, 340 (1989) 35–42.
13 Herr, W., Sturm, R.A., Clerc, R.G., Corcoran, L.M., Baltimore,

D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Finney, M., Ruvkun, G. and Horvitz, H.R., The POU domain: a large conserved region in the mammalian *pit*-I, *oct*-1, *oct*-2, and *Caenorhabditis elegans unc*-86 gene products, *Genes Dev.*, 2 (1988) 1513–1516.

14 Kapiloff, M.S., Farkash, Y., Wegner, M. and Rosenfeld, M.G., Variable effects of phosphorylation of Pit-1 dictated by the DNA response elements, *Science*, 253 (1991) 786–788.

15 Kemler, I., Bucher, E., Seipel, K., Müller-Immerglück, M. and Schaffner, W., Promoters with the octamer DNA motif (ATGCAAAT) can be ubiquitous or cell type-specific depending on binding affinity of the octamer site and Oct-factor concentration, *Nucl. Acids Res.*, 19 (1991) 237–242.

16 Kemp, L.M., Dent, C.L. and Latchman, D.S., Transcriptional repression of HSV immediate-early genes and of octamer containing cellular promoters in neuronal cells is mediated by the octamer DNA element, *Neuron*, 4 (1990) 215–222.

17 Lenardo, M.J., Staudt, L., Robbins, P., Kuang, A., Mulligan, R. and Baltimore, D., Repression of Igh enhancer in teratocarcinoma cells associated with a novel octamer factor, *Science*, 243 (1989) 544–546.

18 Manley, J.L., Fire, A., Cano, A., Sharp, P.A. and Gefter, M.L., DNA dependent transcription of adenovirus genes in a soluble whole cell extract, *Proc. Natl. Acad. Sci. USA*, 77 (1980) 3855–3859.

19 Meijer, D., Graus, A., Kraay, R., Langeveld, A., Mulder, M.P. and Grosveld, G., The octamer binding factor Oct6: cDNA cloning and expression in early embryonic cells, *Nucl. Acids Res.*, 18 (1990) 7357–7365.

20 Quinn, J.P., Variation in the composition of the AP1 complex in PC12 cells following induction by NGF and TPA, *Mol. Cell. Neurosci.*, 2 (1991) 253–258.

21 Quinn, J.P. and Farina, A.R., The autoimmune antigen Ku is enriched on oligonucleotide columns distinct from those containing the octamer binding protein DNA consensus sequence, *FEBS Lett.*, 286 (1991) 225–228.

22 Quinn, J.P., Takimoto, M., Iadarola, M., Holbrook, N. and Levens, D., Distinct factors bind the AP1 consensus sites in Gibbon Ape leukemia virus and simian virus 40 enhancers, *J. Virol.*, 633 (1989) 1737–1742.

23 Revoltella, R., Bertolini, L., Pediconi, M. and Vigneti, E., Specific binding of nerve growth factor (NGF) by murine C1300 neuroblastoma cells, *J. Exp. Med.*, 140 (1974) 437–451.

24 Roberts, S.B., Segil, N. and Heintz, N., Differential phosphorylation of the transcription factor Oct1 during the cell cycle, *Science*, 2 (1991) 1022–1026.

25 Ruvkun, G. and Finney, M., Regulation of transcription and cell identity by POU domain proteins, *Cell*, 64 (1991) 475–478.

26 Schreiber, E., Harshman, K., Kemler, I., Malipiero, U., Schaffner, W. and Fontana, A., Astrocytes and glioblastoma cells express novel octamer-DNA binding proteins distinct from the ubiquitous Oct-1 and B cell type Oct-2 proteins, *Nucl. Acids Res.*, 18 (1990) 5495–5503.

27 Singh, H., Sen, R., Baltimore, D. and Sharp, P.A., A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes, *Nature*, 319 (1986) 154–158.

28 Smith, D.P. and Old, R.W., *Xenopus laevis* Oct-1 does not bind to certain histone H2B gene promoter octamer motifs for which a novel octamer-binding factor has high affinity, *Nucl. Acids Res.*, 19 (1991) 815–821.

29 Staudt, L.M., Clerc, R.G., Singh, H., LeBowitz, J.H., Sharp, P.A. and Baltimore, D., Cloning of a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif, *Science*, 241 (1988) 577–580.

30 Strauss, F. and Varshavsky, A., A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome, *Cell*, 37 (1984) 889–901.

31 Sturm, R.A. and Herr, W., The POU domain is a bipartite DNA-binding structure, *Nature*, 336 (1988) 601–604.

32 Suzuki, N., Rohdewohld, H., Neuman, T., Gruss, P. and Schöler, H.R., Oct-6: a POU transcription factor expressed in embryonal stem cells and in the developing brain, *EMBO J.*, 9 (1990) 3723–3732.

33 Takimoto, M., Quinn, J.P., Farina, A.R., Staudt, L.M. and Levens, D., *fos/jun* and octamer-binding protein interact with a common site in a negative element of the human c-*myc* gene, *J. Biol. Chem.*, 264 (1989) 8992–8999.

34 Tanaka, M. and Herr, W., Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation, *Cell*, 60 (1990) 375–386.

Identification of a neuronal specific DNA-binding protein within the rat preprotachykinin promoter

SUSAN C. MENDELSON AND J.P. QUINN

MRC Brain Metabolism Unit, Royal Edinburgh Hospital, Morningside Park, Edinburgh, EH10 5HF.

The tachykinins are a family of biologically active neuropeptides. They function as neurotransmitters in both the central and peripheral nervous systems where they have a wide but selective distribution (1). The family of tachykinins includes substance P whose precursor is encoded by the preprotachykinin-A gene (PPT) (2).

No cell lines that have been tested express PPT at high levels, either endogenously (3), or when transfected with plasmids containing fragments of the PPT promoter (4), so hampering our understanding of the transcriptional regulation of the PPT promoter. However, a construct linking a 3.5 kilobase region of the rat PPT promoter to a β-Gal reporter gene is expressed in microinjected cultured dorsal root ganglia (Peter Mulderry, personal communication). This demonstrates that this region of the PPT promoter contains *cis* acting elements which will regulate expression of the PPT gene.

We have used an exonuclease mapping assay (5) to explore the mechanisms by which the PPT promoter is regulated. This assay identifies sites of protein/DNA interactions which could be involved in the regulation of gene expression. Multiple and specific DNA interactions are identified within a similar 3.5 kilobase region of the rat PPT promoter to that used in microinjection studies (6). The majority of the interactions observed are present in all cell extracts examined. However, we have identified one site of interaction which occurs 3' of the transcriptional start site and displays neuronal specificity in its tissue distribution. The protein which recognises this site is present in extracts from neuronal tissue (cerebellum, hippocampus and spinal cord) and is also present in the PC12 line which is derived from a rat pheochromocytoma. Binding to this 3' element is not observed in extracts from spleen and the non-neuronal derived cell lines HeLa and 293.

The sequence specificity of the interactions observed within the PPT promoter is suggested by the demonstration of an AP1 complex forming a specific interaction at a previously proposed AP1 consensus sequence (7).

Elements identified by exonuclease analysis of other promoters have proven to have a biological function (5,8,9), suggesting that the sites identified on the PPT promoter are potential *cis* acting elements regulating PPT expression. Further characterisation of these potential regulatory elements will provide a basis to elucidate the mechanism of PPT transcriptional regulation.

We are currently undertaking further analysis of the neuronal-specific interaction with a view to purifying the factor(s) responsible for the interaction.

References

1. Maggio, J.E. (1988). Ann. Rev. Neurosci. **11**, 13-28.
2. Helke, C.J., Krause, J.E., Mantyhl, P.N., Couture, R. & Bannon, M.J. (1990). FASEB J. **4**, 1606-1615.
3. Wood, J.N., Bevan, S.J., Coote, P.R., Dunn, P.M., Harmar, A.J., Hogan, P., Latchman, D.S., Morrison, C. & Rougon, G. (1990). Proc. R. Soc. Lond. B **241**, 187-194.
4. Gilchrist, C.A., Morrison, C.F., Chapman, K.E. & Harmar, A.J. (1991). DNA & Cell Biology **10**, 743-749.
5. Quinn, J.P., Holbrook, N. & Levens, D. (1987). Mol. Cell. Biol. **7**, 2735-2744.
6. Quinn, J.P. (1992). Mol. Cell. Neuro. **3**, (In press).
7. Carter, M.S. & Krause, J.E. (1990). J. Neurosci. **10**, 2203-2214.
8. Takimoto, M., Quinn, J.P., Farina, A.R., Staudt, L.M. & Levens, D. **264**, 8992-8999.
9. Zajac-Kaye, M., Avigan, M., Takimoto, M., Pattuluga, S., Quinn, J., Gelmann, E. & Levens, D. (1988). Curr. Top. Microbiol. Immunol. **141**, 247-252.

Identification of potential regulatory elements within the rat pre-protachykinin A promoter

Susan C. Mendelson and John P. Quinn

MRC Brain Metabolism Unit, Royal Edinburgh Hospital, Morningside Park, Edinburgh, EH10 5HF

The rat preprotachykinin-A gene (rPPT-A) encodes the precursor of the neuropeptide substance P [1]. Substance P is present in specific neuronal subpopulations of both the peripheral and central nervous systems [2]. In particular rPPT-A is expressed at high levels in primary sensory neurons of the dorsal root ganglia (drg) [3].

We wished to investigate the transcriptional regulation of the rPPT-A gene. Deletion studies have demonstrated that 865bp of sequence immediately upstream of the transcriptional start site direct high levels of reporter gene expression when microinjected into adult rat drg neurons in culture. This expression was significantly increased when 500bp of 3' sequence was included in the constructs [4]. This suggests that these regions contain important regulatory elements for controlling rPPT-A gene expression. Exonuclease analysis [5] of the rPPT-A promoter complements these studies by demonstrating the presence of multiple protein/DNA interaction sites both 5' and 3' of the transcriptional start site.

We set out to accurately map sites of protein/DNA interaction within this region (which would be candidate regulatory elements) using DNase1 footprinting analysis. We have thus identified multiple elements located both 5' and 3' of the major transcriptional start site (Fig. 1). The 5' elements include a number of transcription factor binding sites including ones which would be recognised by proteins of the AP1 and CREB families. The 3' elements include one which is predominantly recognised by proteins present in neuronal tissues [6]. Interestingly there are also two G-rich elements, one 5' and the other 3' of the transcriptional start site. There are reports in the literature of G-rich binding proteins which have been shown in one case to be involved in regulating chromatin structure [7], and in another to mediate the effects of EGF [8]. Other binding sites have also been identified whose function is as yet unknown.

Gel retardation analysis has been used to confirm the sequence specificity of individual elements, examine the tissue distribution of proteins recognising the sites and to determine how the proteins are regulated by various extracellular stimuli.

Most of the sites have been footprinted using HeLa nuclear extract, although in some cases rat cerebellum and total brain extract have also been used. Although we have not been able to obtain sufficient material from rat drg for use in footprinting assays some of the identified elements have been shown to be functional in microinjected drg neurons and have formed specific complexes in gel retardation assays using drg extract.

In summary, we have identified several potential regulatory elements within a 1300bp region of the rPPT-A promoter. We are now investigating the function of these elements and plan to mutate individual elements in the context of the promoter to determine their importance in controlling rPPT-A gene expression.

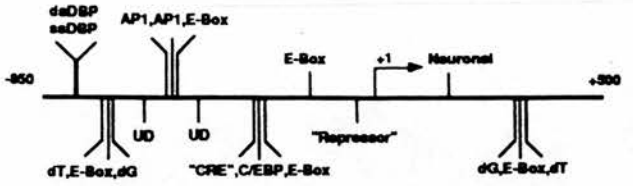


Fig1. Location Of Identified Potential Regulatory Elements On The Rat Pre-protachykinin-A Promoter

A diagrammatic representation of the rPPT-A promoter with identified protein/DNA interaction sites shown. Most of these have been identified using DNase 1 footprinting and further characterised by gel retardation and functional analysis in heterologous reporter constructs. UD = Footprinted elements which we have not yet assigned to a particular transcription factor family (not drawn to scale).

References

1. Helke, C.J., Krause,J.E., Mantyhl, P.N.,Coutre, R & Bannon, M.J. (1990) FASEB J. **4**,1606-1615
2. Maggio, J.E. (1988) Ann. Rev. Neurosci. **11**,13-25
3. Otsuka, M., Konishi, S.,Yanagisawa, M., Tsunoo, A & Akagi, H. (1982) CIBA Foundation Symposium 91: Substance P in the nervous system (R.Porter and M.O'Connor, Eds.) 13-30.Pitman,London
4. Mulderry, P., Chapman, K.E., Lyons,V. & Harmar, A.J. (1993) Mol. Cell. Neurosci. **4**, 164-172
5. Quinn, J.P., Holbrook, N. & Levens, D. (1987) Mol. Cell Biol. **7**,2735-2744
6. Quinn, J.P. (1992) Mol. Cell Neurosci. **3**,11-16
7. Clark, S.P., Lewis, C.D. & Felsenfeld, G. (1990) Nucleic Acids Res. **18**,5119-5126
8. Bachwich, D., Merchant, J. & Brand, J. (1992) Mol. Endocrinol., **6**,1775-1184

The preprotachykinin A promoter interacts with a sequence specific single stranded DNA binding protein

Jane McAllister, Susan Mendelson and John P. Quinn.
MRC Brain Metabolism Unit, Royal Edinburgh Hospital, Morningside, Edinburgh, EH10 5HF.

An element within the preprotachykinin A (PPT) promoter is highly homologous to an element from the type II Na channel promoter (Fig. 1A). This Na channel element has been previously proposed to be common to a number of neuronal genes [1]. We have demonstrated that the PPT element binds a sequence specific DNA binding protein [2]. The protein binds to only one strand of the PPT element and has no specificity for the double stranded DNA species. Gel retardation analysis indicates that the protein is found in both rat neuronal tissue and adult dorsal root ganglia neurons in culture, but not in established tissue culture cell lines. A mechanism for protein binding to the DNA is proposed based on the fact that the region binding the protein is the loop of a larger stem-loop structure in the DNA.

The PPT element can also be aligned with other single stranded DNA elements [3,4] which are known to specifically bind proteins to give a core consensus of CCNGG (Fig. 1B). The PPT element is also highly homologous to the HIV TAR region which, as a single stranded RNA molecule, binds to cellular and HIV viral proteins [5,6].

The initial oligonucleotide used in this study did not contain all the DNA sequence to reconstitute the stem loop structure. When the analysis was repeated using an oligonucleotide which contained the whole of the sequence in the stem loop structure (Fig. 1C) the situation was far more complex than initially thought. The longer oligonucleotide will similarly bind to a single stranded DNA binding protein in rat tissue extract, however, the double stranded element will now also bind a complex in tissue culture cell lines. This binding of a protein to the double stranded element was not seen when rat tissue was used. The protein which binds to the double stranded form of the DNA is adjacent to the binding of the single stranded protein. The binding characteristics of proteins to this region was confirmed by DNase 1 analysis.

Figure 1A

		Position
PPT PROMOTER	TGTG CCtGG GAaGaAGCTGT	-761 to -741
RAT TYPE II Na CHANNEL	TGTGACCAGGAGATGGAGCTGTCG	-60 to -36

Figure 1B

Homology between the PPT element and other SS Nucleic Acid elements which specifically bind to nuclear proteins

		Position
PPT	gTGTG-CCtGG-GAaGaAGCTGT	
PRE (proximal element growth hormone)	gaccGCaGGaG AGCaGTg	-169 to -152
Adipsin	TcTGC	
	cca CCaGGc AAGggGCaGg aggt	-69 to -42
HIV TAR	aG CCtGG G AGCTcTct	
CORE	CCNGG	
PPT	gTGTGCCTGGGAaGaAGCTGT	
HIV TAR	TGaGCCTGGG AGCTcTct	

Figure 1C

PPT sequence of stem/loop structure

5' CCCCTTCTGCCTTC<u>AGGGTGTGCCTGGGAA</u>GAAGCTGTAGGGG

The sequence underlined is the loop of the proposed stem/loop structure

References

1. Maue, R.A., Kraner, S.D., Goodman, R.H. & Mandel, G. (1990). Neuron **4**, 223-231.
2. Quinn, J.P. & McAllister, J. (1993). Nucleic Acids Research **21**, 1637-1641.
3. Wilkison, W.O., Min, H.Y., Claffey, K.P., Satterberg, B.L. & Spiegelman, B.M. (1990). Journal of Biological Chemistry **265**, 477-482.
4. Pan, W.T., Liu, Q. & Bancroft, C. (1990). Journal of Biological Chemistry **265**, 7022-7028.
5. Sheline, C.T., Milocco, L.H. & Jones, K.A. (1991). Genes and Development **5**, 2508-2520.
6. Dingwall, C., Ernberg, I., Gait, M.J., Green, S.M., Heaphy, S., Karn, J., Lowe, A.D., Singh, M. & Skinner, M.A. (1990). EMBO Journal **9**, 4145-4153.

Evolution of enhancer domains within the preprotachykinin promoter

John P. Quinn, Christine Morrison, Jane McAllister & Susan Mendelson. MRC Brain Metabolism Unit, Royal Edinburgh Hospital, Morningside Park, Edinburgh, EH10 5HF.

The rat preprotachykinin-A promoter (PPT) which regulates the gene encoding the neuropeptides substance P and neurokinin A [1] is not, or is only minimally, active in all of the established tissue culture cell lines which have been analysed [2]. To search for transcriptional regulators within the promoter, we have used small fragments from the PPT promoter linked to heterologous reporter constructs. These reporter constructs allow us to determine the ability of individual PPT DNA elements to stimulate transcription. This approach has identified two functional elements within the PPT promoter, both of which are related to AP1 elements. AP1 elements, in general, bind heterodimers of the jun, fos, CREB and ATF families of proteins [3,4], in addition to several less well characterised protein complexes [5] and tissue-specific complexes [6]. These elements within the PPT promoter are at positions -340 and -196 relative to the major start of transcription. The complexes binding to these elements have been confirmed by gel retardation and DNase I analysis. The AP1 element at -340 in HeLa cells will bind the c-jun/c-fos complex whereas the complex binding to the element at -196 is quite distinct from c-jun, c-fos or CREB. The elements are also functionally quite distinct when placed upstream of a heterologous promoter. The sequence of these two PPT AP1-like elements and the AP1 consensus sequence is shown in Figure 1.

We propose that these elements will synergise with one another and other distinct elements within the PPT promoter. DNase 1 footprinting analysis of this region of the PPT promoter has determined many other specific protein/DNA interactions several of which are adjacent to these AP1 elements [7]. Interestingly, there is significant conservation between the regions which are flanking these AP1 elements. This extended homology is shown in Figure 2. These two regions are potentially recognised by members of the same family of transcription factors. As these transcription factors are regulated in response to a multiplicity of stimuli, we propose a mechanism for stimulus induced regulation of the PPT promoter via these two regions outlined in Figure 2.

This extended homology can be found in other AP1 elements which have been associated with enhancer function. This is most clearly illustrated by the AP1 enhancer element of the Gibbon Ape Leukaemia Virus (GALV) enhancer [5,8]. This is outlined in Figure 2. The viral AP1 element is highly homologous to both of the PPT elements we have characterised over an extended region. This extended homology may define a subset of AP1 elements. We view this extended homology as an AP1-like element flanked by two 6bp boxes GTTCCC and CAGTTT which flank an AP1 element. These two boxes were previously noted to be important in retrovirus LTRs as they were conserved in regions associated with enhancer function [8], as is the case amongst the different GALV strains.

Figure 1

```
     TGAGTCA          AP1 consensus
AGCA TGAGTCA ACT      PPT AP1 at -340
AAAT TGcGTCA TTT      PPT AP1 at -196
```

Figure 2

```
GTTCCCagAAAT    AGaATGAGTCA   aCag   CAGTTT     GALV AP1·(SF)

GTTCCC TAAgtccgAAGcATGAGTCAcTT CG  CTCAGTTT     AP1 at -340

cTTCCC TAA     AAt  TGcGTCA TTtCGaaCcCAaTTT     AP1 at -196
```

The GALV enhancer element from the San Francisco strain is aligned with the two PPT AP1 elements.

References

1. Helke, C.J., Krause, J.E., Mantyh, P.N., Couture, R. & Bannon, M.J. (1990). FASEB Journal **4**, 1606-1615.
2. Personal Observations
3. Hai, T. & Curran, T. (1991). Proceedings of the National Academy of Sciences, USA **88**, 3720-3724.
4. Ryseck, R.P. & Bravo, R. (1991). Oncogene **6**, 533-542.
5. Quinn, J.P., Farina, A.R., Gardiner, K., Krutzsch, H. & Levens, D. (1989). Molecular and Cellular Biology **9**, 4713-4721.
6. Andrews, N.C., Erdjument-Bromage, H., Davidson, M.B., Tempst, P. & Orkin, S.H. (1993). Nature **362**, 722-728.
7. Personal Observations
8. Quinn, J.P., Holbrook, N. & Levens, D. (1989) Molecular and Cellular Biology **7**, 2735-2744